

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:56:11 ; Search time 112.99 Seconds  
(without alignments)  
2709.566 Million cell updates/sec

Title: US-09-361-619-11  
Perfect score: 11694  
Sequence: 1 MNHLYKVFKNKATGTFMAVA.....NGSADTOGHVGAAGAGHFH 2314

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_16.\*  
1: sp.archaea.\*  
2: sp.bacteria.\*  
3: sp.fungi.\*  
4: sp.human.\*  
5: sp.invertebrate.\*  
6: sp.mammal.\*  
7: sp.mmc.\*  
8: sp.organelle.\*  
9: sp.phage.\*  
10: sp.plant.\*  
11: sp.rodent.\*  
12: sp.unclassified.\*  
13: sp.vertibrate.\*  
14: sp.virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	1222.5	10.5	2059	2 Q9PD50	Q9pd50 xylella fas
2	1177.5	10.1	2353	2 P71401	P71401 haemophilus
3	1037	8.9	2712	2 Q9F3X5	Q9f3x5 pasteurella
4	743.5	6.4	1190	2 Q9PC04	Q9pc04 xylella fas
5	697.5	6.0	3705	2 Q9F2D8	Q9f2d8 versinia pe
6	661.5	5.7	1107	2 Q9F2D8	Q9f2d8 salmonella
7	659.5	5.6	1098	2 Q48152	Q48152 haemophilus
8	653.5	5.6	1299	2 Q9F3X6	Q9f3x6 pasteurella
9	588	5.0	2586	5 Q9VTK8	Q9vtk8 drosophila
10	578	4.9	2478	2 Q9RL69	Q9rl69 staphylococ
11	578	4.9	2478	2 Q9LCH2	Q9lch2 staphylococ
12	571.5	4.9	2340	2 Q9ZD91	Q9zd91 rickettsia
13	565.5	4.8	2021	2 Q52857	Q52857 rickettsia
14	560	4.8	3029	2 Q55582	Q55582 synecocyst
15	558.5	4.8	2747	2 Q91800	Q91800 aeromonas s
16	552.5	4.7	4919	2 Q9ZHL0	Q9zhl0 haemophilus
17	546	4.7	2514	2 Q9YJ30	Q9yj30 neisseria m
18	536.5	4.6	5627	2 Q91120	Q91120 pseudomonas
19	534	4.6	1557	2 Q9RNI2	Q9rni2 haemophilus

# ALIGNMENTS

RESULT 1

ID	Q9PD50	PRELIMINARY;	PRT;	2059 AA.
AC	Q9PD50;			
DT	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)			
DT	01-MAR-2001 (TREMBlrel. 16, Last annotation update)			
DE	SURFACE PROTEIN.			
GN	XF1529.			
OS	Xylella fastidiosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;			
OC	Xylella.			
OX	NCBI_TaxID=2371;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=9A5C;			
RX	MEDLINE=20365717; PubMed=10910347;			
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,			
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,			
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,			
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Cartaro D.M., Carter H.,			
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,			
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorfy H.,			
RA	Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,			
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,			
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,			
RA	Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,			
RA	Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,			
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,			
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,			
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,			
RA	Mench C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,			
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,			
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,			
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,			
RA	Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,			
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,			
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,			
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Sawasaki H.E.,			
RA	da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,			
RA	de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,			
RA	Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,			

Q91950 pseudomonas  
O18758 sus scrofa  
Q9xc47 rickettsia  
Q9i2m3 pseudomonas  
Q9kxb5 rickettsia  
Q9xcj4 salmonella  
Q9k0t0 neisseria m  
Q9xcq3 salmonella  
Q9zhl3 haemophilus  
Q9kk99 rickettsia  
Q9kxb1 rickettsia  
Q9k031 haemophilus  
Q9jms5 escherichia  
Q9kka5 rickettsia  
Q48028 haemophilus  
Q9kka9 rickettsia  
Q9f0p9 rickettsia  
Q9kka7 rickettsia  
Q9kxb4 rickettsia  
Q9kk98 rickettsia  
Q9kka2 rickettsia  
Q9kka1 rickettsia  
Q9kka3 rickettsia  
Q9fda0 xanthomonas  
Q9kxb0 rickettsia  
P94750 escherichia

20 527 4.5 2147 2 Q91950  
21 520.5 4.5 13288 6 O18758  
22 516 4.4 2106 2 Q9xc47  
23 511 4.4 2468 2 Q9i2m3  
24 509 4.4 1620 2 Q9kxb5  
25 506.5 4.3 2035 2 Q9xcj4  
26 504.5 4.3 2703 2 Q9k0t0  
27 504 4.3 1963 2 Q9xcq3  
28 504 4.3 4152 2 Q9zhl3  
29 492.5 4.2 1604 2 Q9kk99  
30 492.5 4.2 1618 2 Q9kxb1  
31 483 4.1 1536 2 Q48031  
32 482 4.1 1758 2 Q9jms5  
33 475.5 4.1 1616 2 Q9kka5  
34 472 4.0 1477 2 Q48028  
35 470 4.0 1616 2 Q9kka9  
36 469 4.0 1615 2 Q9f0p9  
37 469 4.0 1616 2 Q9kka7  
38 468.5 4.0 1618 2 Q9kxb4  
39 467.5 4.0 1617 2 Q9kk98  
40 465 4.0 1616 2 Q9kka2  
41 465 4.0 1616 2 Q9kka1  
42 465 4.0 1617 2 Q9kka3  
43 464 4.0 1265 2 Q9fda0  
44 463.5 4.0 1616 2 Q9kxb0  
45 463.5 4.0 2349 2 P94750

Thu Sep 13 14:18:21 2001

RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

RT "The genome sequence of the plant pathogen Xylella fastidiosa.";

RL Nature 406:151-159(2000).

DR EMBL: AE003982; AAF84338.1;

SQ SEQUENCE 2059 AA; 204035 MW; EBA6A71B5DD24E10 CRC64;

Query Match	10.5%	Score 1222.5;	DB 2;	Length 2059;
Best Local Similarity	24.08;	Pred. No. 3.6e+43;		
Matches 599; Conservative 368; Mismatches 855; Indels 679; Gaps 115;				
QY	39	GGSPVRLTRVATLAILVIGATLNGSAYAQNNKSIARCTTCNNDNASASH-EASIAIGSL	97	
DB	13	GKEPLHTILFVALSLSLPTTEKANAQVYI--NDGMDGCGORIVNGSASGVERTVA---66		
QY	98	AKAHANQATAIGSK-PDPNQNANOKAGSHAKGKESIAIGDGYLAEGDASIASIGDDLY	156	
DB	67	TOCSEDPVTVTGARFFGSPGTAAEQGAS-----RNLTGGSLYVN---SGQGVNDV-117		
QY	157	LDRNSTNSKYPNCLJSTLTONHTVLRQIRDSNGSQYRRTAAEGHASTAVGA--MAYAKG	214	
DB	118	-----LNKYSIRMGSVITWN-TVA-----GTNAIAGSAQSSAADALKASLATRASG	164	
QY	215	HFANAFCTRTAEGNSLAVGLTAKAEKG-YTIAIGSNAQAINYGALAGADTRVDL-DY	272	
DB	165	ARAIAGAKASADGVDTVLALGSCATAGTCASSIAIGLASAVN-GAVAVGGGALVTVPDG	223	
QY	273	GIALGYSQILNNNNNNNNNNKAYYPEGNSKSKATGNGL-FSIG---SSTIK-RKII	326	
DB	224	AVALLGLNSVA---STGKGLSGYDPKTKTSTDSAAKWTLAASVIGDVSNTLKTROLS	280	
QY	327	NVGAGYEDTAVNVAQLKAVENLAKR--QITTFKGDNGT-----GVKKLGETILT	374	
DB	281	GIAAGTSNTDAVNAQLKAVDEITASRNLTASGANSVAPGSSVLDLKNLTITKA	340	
QY	375	IKGGFTQ-----ADKLTONNIGVVD--NNTCLKVKYLAKNLSGLETST	417	
DB	341	IGSDVQNLNKNKDKVITPLAVGDALLNTDGIAGTGVLSLTGLAI-----TDGPAPVA	394	
QY	418	KNLTASEKTVYVSGNNTAELOSGGLTFTPTTNASTDKTVYGTGDKLFTDNTALEDTIR	477	
DB	395	SGIDAGSKV-----ISHVAAGAVSET-----STDAVNGSQLNAVQVQASQP	435	
QY	478	ITKDKTGFKNKAGTVDENKPYLDKDLKVGNSLTNNGSLTVNNTIGGSN-KOIQVQADGI	536	
DB	436	VT-----FTGNEGAV-----KRLGQSVVLSGESSTAGTYSGNLKSVDVDEAAG-	479	
QY	537	KFADVNVNNAKFCFTTTRITEEIGCFADGADGKVDKPSYLDKQIQVGVKTIKDSGIN	596	
DB	480	-----RIHQLADSPKFCGVNIN-----GG-----KISGVT	506	
QY	597	AGDKISNVKDATDDTDVATYKQLKQVQ-----QADAG	629	
DB	507	AG-----TEETDAVNFSQLKSISTAVDQGTLTASGNSGSKVASGGTVLDLKNPDG	556	
QY	630	ALQSFSTRDEKGOBFTISNLYS-NGNTPNTFETIFAGENGISINDAKKVKVGDPI	688	
DB	557	NLTISKSDSNVVFNLKDFKVDGMTSGT-----TVVNDGVKVGSDVALGTGTLITDG	612	
QY	689	NGLTTPKLTGSKDGKTLQVIEQVAGS---NDTKNIIRGLSPTLPSITNAGGVTRTQG	745	
DB	613	PAVTASGIDAGSK-----VISHVAAGVYSETSTDAVNG-----SQLNAVQVQAS-QP	658	
QY	746	NTITSDEKSKAASIGDILNTGFLKNNSNSVGFVSTYNTVDFIDGNATTAKVYDETQ	805	
DB	659	VFTTNEGAVKRS-----LQGSVVISGESSTAGMYSGNLKSVVDEAAGRIHQLADSPK	713	
QY	806	TSKVYDVNDVDEKTLIELTCDNGKTKNI-----GVKTKT-----TLTPTNANGKATNFSTT	854	
DB	714	FGNVV--INNGKISGVTAGTEETDAVNFSQLKSISTAVDQGTLTASGNSG-----SKV	766	
QY	855	DNDALVNARDIAENLTAKELHTTKGTADTALQTFKVKKGATDDETIIVGKDGQTONGK	914	

DB	767	ASGTVVDLKNKTGCLNT-----ISKSGSDSNVVFNLKDFKVDGM	805	
QY	915	TVNTLKLKGENGTLVAINKDGVTVTGINTQSLKAGDSTTLNKDGLSIKNPASNEQIQVG	974	
DB	806	T-----SGTTVVNN-----DGVKVGSDVALGTGTLT-----832		
QY	975	ADGVKFAKVDKGNSSCTGIDTSRIITKDQIGFTGANGSLDTPPKLTKKLVKGEVEINT	1034	
DB	833	-----TDGPA-----VTAS	841	
QY	1035	GINAGGKRIITNIOGSDITONSNDVTVGGRVYDLKTELESKINSAAKTAQNSLHFEVSAD	1094	
DB	842	GIDAGSKVISHVAAGVYSETSTDAVNG-----SQLNAVQVQ-----878		
QY	1095	OGNIETVSNPYSYDTSKTSQVITFAGENGITTKVANKGVVGVGIDOTKGLTTPKLTGVNN	1154	
DB	879	-----SQPVFTG-----NEGAVKRLSGSVVISGESSTAGTY	911	
QY	1155	NGKGI--VIDSKDQO-NPITGLSNTLANVTNDGAGHALSOGLANDTKTRAA-----SI	1205	
DB	912	SGNLKSVVDEAAGRIHQLADSPKFGNVVINNGKI--SGVTAGTEETDAVNFSQLKSI	969	
QY	1206	GDVLNAGENLQNGEAVDFVSTYDVF--IDGNATTAKVYDDTSKTSKVYVDN---1259		
DB	970	STAVDQGTLTASGANGSKVASGGTVLDLKNKTGDLTISK-----SGDSNDVVFNLKDFK	1024	
QY	1260	-----VDNKTEVTSKKLGVKTTTLTKTSANGNATKFSAADGDALVKASDIATH	1309	
DB	1025	VDGMTSGTTVWVNDGVKVGSDVALG--TTGL--TTANGPAVTASSIDA-----GSKVISH	1075	
QY	1310	LNTLAGDIOTAKGASOASSSASVVDADGNKVLYDSTDKKYQVNDKGOVDKNKEVAKDKL	1369	
DB	1076	V-----AAGAVSETSTDAVNGSQLNAVQVQASQPVTFETGNE-GAV--KRLGQSVV	1123	
QY	1370	VAAQOTPDGTLAQMNVKSVINKE-----QVNDANKQGINEDN-AFIKGLNKAADTKT	1422	
DB	1124	ISGESSTAGTYSGNLKSVDVDEAAGRIHQLADSPKFGNVVINNGKISGVTAGTEED-	1182	
QY	1423	KNAAYTVGDLNVAQOTPLTFAGDTGTAKKGLTETLTKG-----GOTDNTKLTNNI	1474	
DB	1183	---AVNFSQLKSI-----STAVDQGTLTASGANGSKVASGGTVLDLKNKTGML	1227	
QY	1475	GVVAGTQGTFTV--KLAKDLTNLNSVNAAGTRIDEKGISF-----VDANGQAKANTPVL	1527	
DB	1228	TISKSGSDSNVVFNLKDF-KVDGMTSGTTVVNDGVKVGSDVALGTGTLTANGPAVTA	1286	
QY	1528	NGLDLGGKRISNIGA-AVDD--NDAVNFQFNEVAKTVNNLNQNSGASLPEVVTDANG	1584	
DB	1287	SGIDAGSKVISHVAAGAVSETSTDAVNGSQLNAV-----QVQASQPVTF-----1331		
QY	1585	KPINTGDKGPKAKI-----KGADKYYHANANGVPVVDKDGK---PITDADKLANLA	1632	
DB	1332	-----GNEGAVKRLSGQSVVISGESSTAGTYSGNLKSVDVDEAAGTIHLQLADSPKFGNV	1387	
QY	1633	AHKPLDAGHQVVASLGG--NSDAITLTNLIKSTLPDITPNTGNANAGQASLPSLSAAQ	1690	
DB	1388	-----INNGKISGVTAGTEETDAVNFSQLKSISTAVD-----QGTWLT-TASGAN	1431	
QY	1691	QSNAA--VKDVLNCGFNLOTNHNQVDFVKAYDTVFNVTGTDGADITSVRSADGTMSTIV	1748	
DB	1432	GSKVASGGTVLDLKNKTGCLNTISK-----SGDSNDVVFNLSEDLKEKSTIV	1476	
QY	1749	-NTALAATDDGHNVLKAKDGKEYKADDLMPNGSLKAGKASDAKT-PTGLSLVNPNAKG	1806	
DB	1477	GNTQL---DKDG---VKVSSNVLDSNELY-----ITSHISTSVKTLANGESVNVTVN	1526	
QY	1807	G---STGDVALNLLSKAVFKSKDGTITTTTVSSDGIISQCKDNSSITILSKDGLNVGGKVI	1863	
DB	1527	GDGVNIDDVVVVNDL-----GLSIVG--GASLTLS--GINAGSHKI	1563	
QY	1864	SNVGKGTGKTDAAVNOQLNEVRNLLGLG-NAGDNDAGNOV---NIADIKK-DPN---SG	1915	
DB	1564	TNVTAGTETDVAVNFSQLKSVDKGTWLTASGANGSKVSGGTVLDLKNKTGDLNLAISK	1623	





Thu Sep 13 14:18:21 2001

QY 986 -----GNSSTGIDGTSR----- 997  
Db 1113 DGESBETQEVKAGDKVTFKAGKNLKVKSEKDFVSLQDTLGLTSITLAGTANGRND 1172  
QY 998 -----TKDQIGFTGANGSLDTPKPLKYLKLVGEVEITNTGINAGKKTITNIOGSDITQ 1053  
Db 1173 TGTIVKDKGLTITLANGAACT-----DASNGNTISVTKDGISGAKKEITNVKSA--- 1222  
QY 1054 NSNDVTCGRVYDLTELESKINSAAKTAQNSLHEFSVADEQGNHFTVSNPYSSYDTSKT 1113  
Db 1223 -----LKYKDTQ-NTADEQDKEFHA-----AVKN 1247  
QY 1114 SDVITPAGENGITTKV---NKGVVRVIGDQTK-----GLTTP-----KLTGVNNGKGV 1160  
Db 1248 ANEVFVGRNGATVSAKTDNNGKHTVTTDVAEKAVGDLKEDTQKTKLKVNDTIDGNLL 1307  
QY 1161 -IDSKGQNTITGLSN-----TLANVTN-DGAGHALSQG-----LANDTDKTRAASIGDV 1208  
Db 1308 TVDATKASVAKGEFNAVTTDACTAGTNANERGVVGVKSGNGATATETDKKKVATVGDV 1367  
QY 1209 LNAENLQGNCEAVDFVSTYDVTDFIDGNATTAKVYDDT-----SKTSKVYVDVNDK 1263  
Db 1368 AKK-----INDATEFKVENDDSATIDDSPTDGGANDALKAGD 1405  
QY 1264 TIEVTSKKLGVKTTTLTKTSANGNATKFSAADGDALVKASDIATHLNTLAGDIOTAKGA 1323  
Db 1406 TLTAKGNLKVK-----RUGKNITFALAN-DLSVKSATVSKLS----- 1444  
QY 1324 SOASSASYVDADGNKVIYDSTDKKYQVNDKQVDKNEKVAKDLVACQAPDGTLAGM 1383  
Db 1445 -----LGTNGKNVNTSDTKGLNFAKD-----SKTGDANIHLNGTASTLTDLLNS 1491  
QY 1384 NVKSVINKEQVNDANKKGINEDNAF-----IRGLENAAKDTKTKNAAVTVGDLNAVOT 1438  
Db 1492 GATTNLGGNGITDNEKKRAASVDVNLGNVGRVGPASANNQVENI-----DFVATYDT 1546  
QY 1439 PLTFAGDTGTTAKKLGELFTLTKGGQTDNTKLTNNIG-----VWAGTDG---FTVKLAKDLT 1492  
Db 1547 VDFVSGDKDTT-----SVTVE---SKDNGKRTKVIKAKTSVVKDNGKLTCKELKDN 1598  
QY 1493 NLNSVNAVGTIDEKISFVANDGQAKANTPVLISANGLDL---GGKRISNIGRAVDDNDA 1549  
Db 1599 N-----NGVTTETDGRDEGNLVTAKAVIDAVNKAQVKTGTGANGQND-- 1644  
QY 1550 VNEKQFNEVAKTVNNLNOSNGASLPFVVTDPANGKPTNGDKPQKAIGA-DGKYIHA 1608  
Db 1645 -----FATVASGTN-----VTFADG---NGTAEVTKANDGSITVKYNK 1681  
QY 1609 NANGVPYDKGKPTTDADKLANLAHKG---PLDAGHQVVASLGSNSDAITLNIKSTLP 1665  
Db 1682 VADGLKLDGD-KIVADTTVLT--VADGKVTAPNNGDKKFDVDSGLADALNKLSTAT-- 1736  
QY 1666 QIDTPNTGNAGQAQSLPSLSAAQSNASVYKDLVNLNGENLOTNHNQVDF-----VK 1718  
Db 1737 -----AGKEGTGEVD--PANSAGQEVKAGD-KVTFKAGDNLKIKGSGKDFYSLKKEK 1787  
QY 1719 AYDTVNEFN---GTGADITSVRSAGDTMSNITV---NTALAATDDGNVLIAKAGKGYK 1772  
Db 1788 DLTSVEFKDANGTGSESTKI-----TKDGLTITPANGAGAAGANTANTISVTKDG---- 1838  
QY 1773 ADDLMPNGLKAGKASADKPTTGLSLVNPNAKSGTGDVAVALNNLSKAFVKS-----K 1826  
Db 1839 -----ISAGNKA-----VTNVVSGLKKEFGDGHTLANGTADFEKHYDNAYK 1879  
QY 1827 DGTFTTTVSSDGISIQGKDNSSITLSKDLNVG-----GKVIS---NVGKGTDDTAANY 1878  
Db 1880 DLT-----NLDEKGADNNTPVADNTAATVGLRLGLWVVSADKTTGEPNOEYN-AQV 1930  
QY 1879 QOLNEVRNLLGLG-NAGNDNADGNQVNIADIKDP-----NSGSSSNRTVIRKAGTV 1928  
Db 1931 RNANEVKKSGNGINVSCKTLNGTRITVITFELAKGEVVKSNFETVKYKADGSETNLVKVGD 1990

QY 1929 LGKGKND--TEKATGGVOGVVDKDNANGDLSNVWVKTKDGSKKALLATYNAAGOTN 1986  
Db 1991 YYSKEDIDPATSPKMTG---KTEKYKVENGV-----VSANGSKTEVTLTNKSG--- 2037  
QY 1987 YLTNNPAEADIRINEOGIRFFHVNDQNGPVPVQGRNGIDSSASGKHSVAIGFOAK----- 2041  
Db 2038 YVTGN--QVADAIKSGFEL-----GLADAAEAKAFASAKDKOLSKD 2079  
QY 2042 -----ADGEAAVAIGQTOAGNOGTAIGDQASIAIGTGNVVTGKHSGAIGDPS 2095  
Db 2080 KAETVNAHDKVRFANGLNTKV---SAATVESTDANGDK-----VTT-----T 2118  
QY 2096 TVKADNSYSVGNNOQFIDATQTDVFGV---GNNTVTT-----ESNSVALGNSAISAG- 2145  
Db 2119 FVKTD-----VELPLTQIYNTDANGNKIVKADGKWTVELNADGTASNKVETLGN 2167  
QY 2146 --THAGTQAKKSDGTAGITTTTAGATGTVKGFAGOTAVGAVSVG-----ASGAE 2191  
Db 2168 VDANGKKVYKVTENGADKWKYTNADGAADTKGEVSNKYSTDEKHVVRVLDPNQNSGK 2227  
QY 2192 RRIQNAAGEVSATSTDAVNGSOLYKATQGTAN--ATNELDRIHONENKANAGISSAM 2248  
Db 2228 VVIDNVANGEISATSTDAINGSQLYAVAKVTNLGAGQVNNLEGVNKVGRADAGTASAL 2287  
QY 2249 AMASMPQAVIPGRSVMTGGIATHNGSGAVAVGLSKLSKDSQGVFKINGSADTQGHVGA 2308  
Db 2288 AASQLPQATMPKSMVAIAGSYQSQONGLAIGVSRISDNGKVIILRLSGTTNSOGKTVAA 2347  
QY 2309 GAGFHF 2314  
Db 2348 GVGQW 2353  
RESULT 3  
ID Q9F3X5 PRELIMINARY; PRT: 2712 AA.  
AC Q9F3X5;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE MAPB PROTEIN.  
DN MAPB.  
GN Pasteurella multocida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Pasteurella.  
OX NCBI\_TaxID=747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PM70;  
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;  
RT "Evolutionary origins of the autotransporter proteins."  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ277636; CAC14203.1; --  
SQ SEQUENCE 2712 AA; 276154 MW; 3F5579D6F32FFA3D CRC64;  
Query Match 8.9%; Score 1037; DB 2; Length 2712;  
Best Local Similarity 23.1%; Pred. No. 3.1e-35;  
Matches 636; Conservative 312; Mismatches 935; Indels 868; Gaps 132;  
QY 12 ATGTFMAVAECAKSHSGSSSTAGQVSSPVRLTRVATL--AILVIGATINGSAYQN 69  
Db 374 ASDNSIALGNAARTNKDSDSIAIGSAETKAAHSLAVGTTSKALAEALALGKLAEGK 433  
QY 70 NSKTAFGVT-----GNNDNASASNEASIAIGSLAKAHANQAIAG-----GSKP 113  
Db 434 TSSVAMGNTSKADGNSVAVGNTSQTLQSN--TIAIGSSAIANPERTISIGLNACKGQEA 491  
QY 114 DPRNOAANQ-----KASHAKRESTAIGDVLAE--GDASIAIGSDDLYLDRNSTNSKY 166  
Db 492 DATGTKHSQLNIGNSGEGVIGQLNIGIGHAGTNVVGKHNIALGT---YAGTNLKNSEE 548  
QY 167 PNLGLSTLIQNTVLRQIRDSNGSQYRRTAAEGHASTAVGMAYAKGHFANNAFGRSTA 226



Thu Sep 13 14:18:21 2001

Query Match 6.48; Score 743.5; DB 2; Length 1190;  
Best Local Similarity 24.18; Pred. No. 2e-23;  
Matches 352; Conservative 188; Mismatches 481; Indels 441; Gaps 67;  
QY 1010 GSLDTTKPHLTDKLKVGE---VEITNTGINAGGKKTITNQSDITONSNDVATGGR-- 1063  
Db 13 GGIEGVDLP-LFLPKYKIQSLOHAWMTSA--ASSKK-----CTQPRRNNAMATAKRS 62  
QY 1064 -VYD-----LKELESKINSAAKTA-----ONSLEHFSVADEQGNHFTVSNPY 1105  
Db 63 KLHRRRLHLVLLTLAASTGYTKVAAQVYVNSDSTENCVEILGDSSTQTSFIHSASNDK 122  
QY 1106 SSYDTSKTSDVITFAGEN---GITTKVKNVGVVRVIDQTKGLT-TPKL-TVCNNNGK--- 1157  
Db 123 CKPDETQTEVSLFYDYRNVLGSLVYNEG--KLGLVDISGATYSMRGLSITATNNGSAGI 180  
QY 1158 -GIVIDSKDQONTITGLSNTLANVTNDG--AGHALSQGLANDTKTRAASIGDVLNAGFN 1214  
Db 181 DSIATIGSGGSKT-----DONTSGATVAQGL-----RSIALGTTARS--- 217  
QY 1215 LOGNCEAVDFSTYDTVDYDFIDGNATTAKVYDDTSTKTSKVYVDVNDVNDKTIETVSDKILG 1274  
Db 218 -----QSQDAISIGTGASTGTGNEAF----- 237  
QY 1275 VKTTLTKTSANGNATKESAAQDGLVKASDIATHLNLTLAGDITQAKGASQASSASYVD 1334  
Db 238 -----AIGNGALTSIANGIALGASSVTTT-----GGVALGQGSAAATASGI--- 279  
QY 1335 ADGNKVIYDSTDKYQVNDKQGVKNKEVAKDKLVAQAOTPDGTLAQMNVKSVYNKEQV 1394  
Db 280 -----TCYDPVTKS-----TSTLSMTSMRSTLGAWSI 306  
QY 1395 NDANKKOGINEDNAFTKGLENAAKTKTKNAATVYDGLNVAQT-----PLTFAGDTGTT 1449  
Db 307 GNTSS---TSQTRQLTGLAAGRSDDT---AVNVAQLKLLAESYGGGWNLTASGSSN 359  
QY 1450 AKKIGETLITKGGQTDITNKLTNDNIGVAGTGC--FTVKLAKDLNLNLSNAGGTTRIDEK 1507  
Db 360 V-ALGESVDLKN-----SDGNLLITKTDNDVTFNLATL-KVDSLTTGNTAMTDT 409  
QY 1508 GISP-----VDANGOAKANTPVLSANGLDLGGKRISNIGANVDNDVNFKOFNEVAKTV 1562  
Db 410 GVTGKRVTLTOSTGLVIAEGSPVSISSGINAAGQKIMNVGTGTADTDAVNFQGLQAVSDTA 469  
QY 1563 NNLNNQSNQSGASLPVVTDANGKPIGTGDKPKQKAIKGADKYHYHANANGVPVDKDGKPI 1622  
Db 470 SKGWNLLASGTN-----SSNVAPGASVD-----LAKNTDGNLLITKAIGI----- 508  
QY 1623 TDADKLANLAHAGKPLDAGHGVVAVSLGNSDAITLITNKTLPQIDTPNTGNA-NAGQAQ 1681  
Db 509 --NDVTFNLAT---ALEADSLTTGNTAMTDTGTVGS-----NVTLSGLVITDGPVS 557  
QY 1682 SLPSLSAAQO---SNRAQVYKVLNVLNFGNLOTNHNQVDFVKAYDTVNF-----VNGTGADI 1733  
Db 558 TSQGISAGNOKITNVAAGTADT-----DAVNFSQLQAVSSTASKG 597  
QY 1734 TSVSADGTWISNTVTNTALAAATDDGCVLI-----KAKDGKEYKADDLMPNGLSKAGKAS 1789  
Db 598 WNLASGANSNVVPGESVDLKNSDGNLLITKTDNDVTFNLATLAKVD-SLTTGNT-- 654  
QY 1790 DAKPTPTGLSLVNPAGSGTGVDAVNLNLSKAVFKSKDGTNTTTSVSSDGIISQGDKNSSI 1849  
Db 655 -AMTDTGVT-VGSNVTLGSTG-----LVITDGPVS----- 682  
QY 1850 TLSDKGLNAGVKVTSNVYKGTGKTDAAVQVQLNEVR-----NLGLG-NAGN----- 1895  
Db 683 --TSSGISAGNOKITNVAAGTADTDAVNFQGLQAVSSTASKGWNLLASGANSNVAPGES 740  
QY 1896 ---DNADGNVNIADIKKDPNNGS-----SSNRTVTKAGTVILGG 1931  
Db 741 VDLKNTDGNIV-----ISKESGNDVLFNLSSSLKDLKLTGVTVMVTNTGTVGSGVTLGS 796

QY 1968 KDGSKALLATYNAAGTNTYLTNNPAAIDRINEQGIREFHVDNGOBPPVQGRNGIDSS 2027  
Db 2443 KDGKDAVAISGKGVG-TIGLT-CPAGA-DGNANAI--IGVND-----VKGLDNGDK 2492  
QY 2028 ASGKHSUVAIQKADGAEAAVAGROTCAGNQSTAIQDQNAQAQTGDSIAIGTGNVYTKG- 2086  
Db 2493 -DGNKTRIV-TKPNGBEE-----EQVATNDGLVFG--ADKQTEHKAKLGTVKYKGD 2543  
QY 2087 ---HSCAIGDPSTVKAADRSYSVGNNOFIDATQTDVFGVGNNTVTFSNVALGSNSAIS 2143  
Db 2544 KNTEVEAGDTIRVLKDNIDVKGIN-----VTENLTVEKAGKINMNV--- 2588  
QY 2144 AGTHAGTQAKSDGTAGTTAGTGTGKGFAGGTAVGAVSVGASGAEIRIQNVAAGEVS 2203  
Db 2589 -----IDGVADEGVN 2598  
QY 2204 ATSTDAVNGSOLYKATQGI---ANATNELDRIHQENKANAGISSAMAMSPQAYIPG 2260  
Db 2599 ATSKQAVNGSOLHVKQOVQNNQATAIKGLDHINKVDKDLRAGIAGATAVAFLOPNEAG 2658  
QY 2261 RSMVTGGIATHNGGAVAGVAVKSLSDNGQWFKINGSADTQGHV--GAAGV 2309  
Db 2659 KSVLSVGVSGSYRESATAVGARNSDNNKISIKLGGGMNSRGDVNFGSGIG 2709

RESULT 4  
Q9PC04 PRELIMINARY; PRT; 1190 AA.  
ID Q9PC04  
AC Q9PC04  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE SURFACE PROTEIN.  
GN Xf1981.  
OS Xylella fastidiosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
OC Xylella  
OX NCBI\_TaxID=2371;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9A5C;  
RX MEDLINE=20365717; PubMed=10910347;  
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
Alvarenga R., Alves L.M.C., Arya J.E., Bala G.S., Baptista C.S.,  
Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,  
Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
Facincani A.E., Ferreira A.J.S., Ferreira V.C.A., Ferrio J.A.,  
Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,  
Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marinho C.L.,  
Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
Vallada H., van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";  
RL Nature 406:151-159(2000).  
DR EMBL: AE004017; AAF84783.1;  
SEQUENCE 1190 AA; 118446 MW; 756741B0C8D787CC CRC64;



[illegible]

```
QY 1017 PHLTKDKLVGEVEITNTGINAGGKKTITNIQSODITONSNDVAVTGRVYDLTKTELESKIN 1076
Db 97 -----NGSKSITNLAAGTTLAADSTDAVNGSOLFDTNEKVD--- 132
QY 1077 SAAKTAQNSLHIEFSVADEQGNHFTVNPYSSYDTSKTSVDITFAGENGITTKVNGVVRV 1136
Db 133 -----ON-----TADITT-----NTNSINQNTDI 152
QY 1137 GIDOTKGLTTPKLTGVNNGKGIVIDSKDQNTITGLSNTLANVNDGAGHALSOGGLAND 1196
Db 153 ATWTT-----SINLSNSVTTLTD-----ALLWDAASG 181
QY 1197 T-DKTRAASIGOVLN--AGFNLOGNGEAVDFVSTYDTVFIDGNATKATKYVDTSKTSK 1253
Db 182 TFSASRNGSASKITNLAAGTTLAADSTDAVNGSOLFDTNEKVDQN-----TADITTNTNS 235
QY 1254 VYDVVNDNKTIETVDKGLGVKTTTLTKTSA--NGNATKFSAADGDALVKASDIATHLN 1311
Db 236 I--NQNTDIATWTTTINNLSNSVTTLTDALLWDAADSGTFSASRNGSASKITNLA--G 291
QY 1312 TLAGDIQTAKGASQASSASYVDADGNKVITYDTRKKYQVNDKQVKNKEVAKDKLVA 1371
Db 292 TLAADSTDAVNGSQ-----LYETNQK--VDQNTSAID--- 322
QY 1372 QAOTPDGTLAQMNVKSVINKEQVNDANKKOGINE-DNAFTKGLNENAAKDKTKTNAAVTVG 1430
Db 323 -----INTSITNLSNDLSNWNETTNSF-----SASHGSSTTNKI--- 356
QY 1431 DLNAVAQTPLTFAGDTGTTAKLGEILT--IKGGQT-DTNKLTDDNIGVVAGTDGTVKL 1487
Db 357 -----TNVAAGEUSEESTDAVNGSOLFETNEKVDQNTDIAANTNITON 401
QY 1488 AKDLTNLNS-----VNAAGGTTRIDEKGISFVDANGQAKANTPVLSSANGLIDLGG--KRISNIG 1541
Db 402 STAIENLNTSVSDINTSITGLTDNALLWDEDTGAFSANH-----GGSTSKITNVA 451
QY 1542 AAV---DDNDAVNFKQFNEVAKTVNNLNOSNGSGASLPFVVTDANKPIN-GTDGPKPKA 1597
Db 452 AGALSEDSTDAVNGSOLYETNQVD-----QNTSA-----IADINTSITNLTGTD-----A 496
QY 1598 IKGADGKYIHANAGVPVVDKGRPIITDADKLANLAHGRFLDAGHQVVASLGSNSDAITL 1657
Db 497 LSWDDDEGAFSAGHT-----SGTNKITNVA--GE-----IAS--DSTDAING 536
QY 1658 TNIKSTLPQIDTPTNTGNANAGQAQSLPSLSAAQO-----SNAASVKVDVLNVGFNLQTNHN 1712
Db 537 SOLYETNMLISQYN-----ESISOLAGDTSEYIITENGTVK-----YIRTDN 580
QY 1713 QVDFVKAYDTVNVFNGTGADIT-----SVRSADGTMSNTITVNTALA 1753
Db 581 GLEGDQAYATGNGATAVGYDAVASGAGCLALGQNSSSSIEGSLALGSGTNSRAITTGIR 640
QY 1754 ATD--DDGNVLIRAKDGKFKADDLPNGSLKAGKSASDAKTPGLSLVNPAGKSTG- 1810
Db 641 ETSATSDGVI-----GNTTDRELLGALSIGTGESYRQIT-----NVADGSEAQ 686
QY 1811 DAVALLNLSKAVKSKDGTITTTVSSDGISIQGKDNSSITPLSKDGLNVGKVISNVGKGT 1870
Db 687 DAVTVRQLQNAI-----GAVTTT-PTKYHYHANSTEEDSLAVGTDLSAMGAKTIVNADAG- 739
QY 1871 KDTDAANVQQLNEVRNLLGLGNAGNDNADGNQVNIADIKDPNCGSSNSTVTKAGTVLG 1930
Db 740 -----IGICLNPLVMADAINGIAIG-----SNARANHANSIAM 772
QY 1931 KGNNDTEKLATGVQGVQDKDGNGLDLSNVVVKTKDKGSKALLATYNAAGQ----- 1984
Db 773 GNGSQTRGAQDTYAYNMDTPONSNGEFS-----VGSEDGQRO-----ITNVAAGSADTDAV 825
QY 1985 -----TNYLTNPAPADIRINEQGRFRFHVNDGNOEPVVQGRNGIDSSASGKHSVAIG 2037
Db 826 NVGOLKVTDAQVSRNTQSTINLTQVSNL-----DTRVTNIENGIGDIVTGTSTKY----- 876
```

```
QY 2038 FQAKADGERAAVALGROTAQGNQSIATGDNAAQATGDOSIAIGTGNVTVGKHSIGAIDPSTV 2097
Db 877 FKNTTDGA-----DANAQAGDSVAICSGSIAAENSVALGTNSVA 916
QY 2098 KADNSYSVGNNOFIDATQTDVFCVGNNTVTSNSVALGNSAISAGTHAGTOAKKSDG 2157
Db 917 DEANTVSVGSSTO-----QRRITNVAAGVNTDAVNA-----OLKASE- 955
QY 2158 TAGTITTTAGFATGVKGFAGQTAVCASVSG-ASGAERRIQNVAAAGEVSATSTDAVNGSOLY 2216
Db 956 -----AGSVRYETNADGSVNVLNLGDSGGTTRIGNVSA---AVNTDVAVNAQLK 1005
QY 2217 KATOGTIANATN-----ELDHRIHQENKANAGISAMAMASMPQAYIPGRSKVGTGIATHN 2272
Db 1006 RSYEEANTYTDQKMGEMNSKIKGINKMSGIASAMAGLPOAYAPCANNTSIAGGTFN 1065
QY 2273 GOGAVAVGLSKLSDNGOWFKINGSADTQGHVGAAGVAGGPHF 2314
Db 1066 GESAVALGVSMSVSESGWVYKLOGTSNSQGDYSAAIAGAGFQW 1107
RESULT 7
Q48152 PRELIMINARY: PRT: 1098 AA.
AC Q48152;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
DE ADHESIN (HIA).
GN HIA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NONTIPERABLE STRAIN 11;
RX MEDLINE=96332658; PubMed=8730864;
RA Bareilakamp S.J., St Gene J.W. III;
RT "Identification of a second family of high-molecular-weight adhesion
RT proteins expressed by non-typable Haemophilus influenzae.";
RL Mol. Microbiol. 19,1215-1223(1996).
DR EMBL: U38617; AAC34721.1; -.
SQ SEQUENCE 1098 AA; 114100 MW; D977335A89F7333D CRC64;
Query Match 5.6%; Score 659.5; DB 2; Length 1098;
Best Local Similarity 22.9%; Pred. No. 5.9e-20;
Matches 323; Conservative 168; Mismatches 457; Indels 461; Gaps 55;
QY 1008 ANGSLDPTTKPLTKDKLVK---GEVEITNTGINAGGKKTITNIQSODITON----- 1054
Db 49 ANNTPTV-----NKLKAYGDANFNFTNNSIADAQVQVQVQVAYKGLNLNENKNASDKLV 102
QY 1055 -SDAVTGGRYDLTKTELESKINSAAKTAQNSLHIEFSVADEQGNHFTVNPYSSYDTSKT 1113
Db 103 EDNTAATAVGRLKGLWVLSKNGTNRKESQOVKHADEVLPFEGKGVQVT---STSENGKH 159
QY 1114 SDVITFAGENGITTKVNGGVVRVGVGIDQTKG-LTTPKLTIV-CNNNGKGVIVDS--KDGQNT 1169
Db 160 TIYFALAKDLGVKTATVSDTLTIGGGAAGATTTTPKVNVTSTTDLGLKPAKDAAGAGDGT 219
QY 1170 I--TGLSNTLANVNDGAGHALSOGGLANDTKTRAASIGDVNLNAGFNILQ-----GNG 1219
Db 220 VHLNGIGSTLTD-TLVGSPATHIDGGDQSTHYTRAASIKDVLNAGWIKGVKAGSTTQGS 278
QY 1220 EAVDFVSTYDTVFIDGNATKATKYVDTSKTSKVYDVVNDNKTIETVDKGLGVKTTT 1279
Db 279 ENVDVFVHTYDTVEFLSADTETTTVTVDKSKNGKRTKVIKAKTSVIREKDGKLTGKANK 338
QY 1280 LTKTSANGNATKFSAADGDALVKASDI-----ATHLNTLAGDIQT-AGKASQA 1326
Db 339 ETNKGVDGANATE-DADEGKGLVTAQKVIDAVNKTGWRIKTTDANGONGDFATVASGTNVT 397
```



RESULT 8  
 Q9F3X6 PRELIMINARY; PRT: 1299 AA.  
 ID Q9F3X6  
 AC Q9F3X6;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE MAPA PROTEIN.  
 GN MAPA.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;  
 OC Pasteurella.  
 CC Pasteurella-747;  
 OX NCBI\_TaxID=747;  
 RN [1]  
 RS SEQUENCE FROM N.A.  
 RP STRAIN=PM70;  
 RC Henderson I.R., Nataro J.P., Cappello R., Stein C.;  
 RA "Evolutionary origins of the autotransporter proteins."  
 RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AJ277635; CAC14202.1;  
 DR EMBL: AJ277635; CAC14202.1;  
 SQ SEQUENCE 1299 AA; 130963 MW; 8BCCE0EB66CDB428 CRC64;

Query Match 5.6%; Score 653.5; DB 2; Length 1299;  
 Best Local Similarity 23.3%; Pred. No. 1.3e-19;  
 Matches 402; Conservative 207; Mismatches 554; Indels 563; Gaps 84

Qy	699	GSDRDXGTQLVIEQVAGSNDTRKNIKGLSPTLPS---	ITNAGGVRTTE-QGNTITSDEDK 754
Db	27	GKSASGKSALV-----NSVSGSFLLIAASVVLGSGQVNAAEVTGNTGVSGDDK 75	
Qy	755	SKAASIGDILNTGPNLKNNSVGFVSTYNTVDFIDGNATTAKVYTDETNOTSKVTYDVN 814	
Db	76	-----YCFYNASQ-----SVIGGDATTK--TTDKTD-----N 101	
Qy	815	VDEKTIEL---TGONGKTN-KIGVKTTTLTTNANGKATNFSTNDALVNAKOENLN 870	
Db	102	KPAKSWGIFGATNDGETINVAIGAK-----SKSKAASIAIGDNAKALDNOAIAIGON 154	
Qy	871	TLA-KELHTTKTADTALQTFKVKKDGATDEITVKGDKGTQNGKTVNTL-----K 920	
Db	155	ATANSWDWISIGROAGAEOT-EVSAEG-----RNIAIGDGALKKRGQVNNNIALGTSAGRD 209	
Qy	921	LKGENGLVAT--NKDGTVTGINTQSGLKAGDS-----TTLNKGDLGSLKNPAS--- 967	
Db	210	LAGTHNVLMGTYNDAEV-----RSALTAGSTTKEINAKETIDNKRYTEASNTVALGT 263	
Qy	968	-----NEQIQVGADGVKPAKVDKNSSTGIDGTSRTKDQIGFTGANGSLDITTKPPLTK 1021	
Db	264	RALATQLAAVAIGQAKAF-----GNQSVAVNGNKGKAGQTATAGSNA-----HAT- 310	
Qy	1022	DLKQVEVEITNTGINAGKKITNTQSGDIT-----QNSNDAVTGGRRYVDLKTETLES 1073	
Db	311	-----GSSSIAIGTVNGNDTARTLASDNFTIAMLSAQATKSDAIAVGR----- 355	
Qy	1074	KINSAAKTAQNSLHEFSVADEGNHFTV-----SNPYSSYDTSKTSDDVITFAGENGITKV 1129	
Db	356	-----NAKAA--GINTSVISIGNAGVTATADAOSDPNKKLEPA--TDVFIGNKAGYKSNQ 407	
Qy	1130	NGVVRVCIQDTKGLT--TPKLTVTGNNNGK-----GVIDSKDQONTITGLSNTLANVTN 1182	
Db	408	NRMQVSLGKDSGEVGVGTENVNTIGNSAGKNTKGTNVAISSRAGQ--VEGHIDNFAA--- 462	
Qy	1183	DGAGHALSQGLANDTDKTRAASIGDVLNAGFNLOGNGEAVDFVSTYDVFIDGNATTAK 1242	
Db	463	-----LEAQNKGK-----DNIAIGK 480	
Qy	1243	VYDDTSKTSKVVDVNVNDKTI-----EVTSDKKLGKVTTLTKTSANGNATKFSAADGD 1298	
Db	481	HAGRSADPNTK---LNINN-TISLGRSVSLKNFGI-----AQGNKAK---TDG- 522	
Qy	1299	ALVKASDIATHLNTLAGDIQTKAGSQAASSASVVDADGNKVIYDSTDKKYYQVNDK--G 1356	

Db 523 ----LASIAIGRNAEAVGGETANIAIGDSASA---DASGAIVLGTAKAQSSTLVGKKY 575  
QY 1357 QVDKNEKAVKDLVAQA-----QTPD-----GTLAQMNVKSVINKEQVNDANKKOCINE 1405  
Db 576 AYSIAVIGTEAKAIAQAAPAGRNENPKDAIAIGTKAEHAYSTI-----619  
QY 1406 DNAFIKGLENAKDKTKNAVTVG--DLNAVAQTPTTFAGDTGTTAKKLGELTITKGGOT 1464  
Db 620 -----ALGFAGK--SDTKAQAVSNGSNKAGYQALAFGSEAKTT--ENAGSSIAF--GTRKA 670  
QY 1465 DTNKLTDNIGVYAGTGDGTFLAKDLTLNLSNVNAGTTRIDEGISFVDANGQAKANTPV 1524  
Db 671 QTRASASIAIGMAET--GFD-----GGQALD--GSDAVALGREAKAK---709  
QY 1525 LSANGLDLGGKRISNIGCAAVDDNDVNFQFNEVA-----KTVNNLNQNSGASLPF 1577  
Db 710 -RONALAFYK-----AVADHKDAVALGAGAEATAEAGTNEATYNEFYSGFAGI---758  
QY 1578 VYTDANGKPIGTGPKQKAIKAGDKGYHANANGVVPVDKGPITPDADKLANLAHAKP 1637  
Db 759 -----KPI-----ATVSVGKKDAERTITNVAAGRDKTSDAIN-----792  
QY 1638 LDAGHOVASLGSNSDAITLTNKTSLTPQIDPTPNTGNANAGQAQSLPSLSAAQSNAAV 1697  
Db 793 ---GSOLYLALN-----ALGNVGNFL-----VTNLVGGDA-----AIVKEGDEAGT 830  
QY 1698 KDVNLVGNFNLQTNHNDVFNKAYDVTNFVNGTG--ADITSVRSADGTMSNITVTNATAATD 1756  
Db 831 LTPMSNIG---GTCKGTI-----HDAEAVNNTAKASKTTVKEGD-----NITVTEAAADG 878  
QY 1757 DGNVLIIKADGKYKADDLMPNGSLKAKSASDAKTPGLSLVNPNGKGTGDAVALN 1816  
Db 879 SRTYTVATKDKVFE-----DSVAVAGTKIDA-----NGLTFVD-----911  
QY 1817 NLSKAVFKSKDGTITTVSSDGSISIOGKNSITLSKDLGNLVGKKVISNVGKG--TKDPT- 1873  
Db 912 -----DQGT-----KIDNTPSTSKTGIDAGNKVTNVQNGNIAKDSK 948  
QY 1874 DAANVOOL-----NEVRNLLGLNAGDNADGNQVNIADIKKDPNSGSSNRTVIRAKTVL 1929  
Db 949 DAVNGQLFAQEGEVKNIG-----GDTYINPETEYAN-----TNI 985  
QY 1930 GKGKGNNDTEKLATGGVGVGVDKGNAGDLSNVYKTKDGSKKALLATYNAGQTNILT 1989  
Db 986 GGTGASTI-----DEAIKAVNTTAKAA-----1007  
QY 1990 NNPAEAI DRINEGIRFFHVNDGNQEPVYVQGRNGIDSSASGRHSVAIGFQAKDGEAANA 2049  
Db 1008 -----KTEVVQGENIVVTSAPGAN-----1026  
QY 2050 IGRQTQAGNOSTAIGDNAONTGDOSTAIGTGNVVTCKHSGAIGDPS--TVKADNSVSGN 2107  
Db 1027 -----GNTVYVATAKEVTFDKTI-----VGSVVTDKNTDITGLSNKLTGDNFAKNR 1076  
QY 2108 --NNQFIDATQTDVGV--GN-----NITVTESNSVALGS--NSAISAGHTAQAK--2154  
Db 1077 AASEEQLNATQTNLATLLGNAQNTNGNVAMTDIGTGKNNINDAIKASRNEVKQCKNMV 1136  
QY 2155 ---SDGTAGTTTAGTGTGVKFGAGTAVGAVSGVSGAERRIQNVAAGEVSATSTDVN 2211  
Db 1137 VPTTGANGOTIYEVATAKVAE--DEYKVGGITIDAT--TNKISGIAKGDISNSTDAVN 1193  
QY 2212 GSOLYKATOCIANA---TNELDRH IHNENKANAGATSSAMAMASMPQAYIPGRSVMVTGGI 2268  
Db 1194 GSOLYELQOKIAKSGDNYNLLNRINKVOKDLRAGTAGANAAAGLPOAYIPGKSMVAVAA 1253  
QY 2269 ATHNGQAVAVGLSKLSDNGQWFKINGSADTQGHYGAAGVAGFHF 2314  
Db 1254 GTYKQGNALAGMSRISDNGKVKIILKTGNTNSRGDFGASIGAGYQW 1299

RESULT 9

Q9VTK8  
ID Q9VTK8; PRELIMINARY; PRT; 2586 AA.  
AC Q9VTK8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAY-2001 (TREMBLrel. 16, Last annotation update)  
DE CG18331 PROTEIN.  
GN CG18331.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=107311132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abriel J.F., Abayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostler C., Gabrielian A.E., Gar N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ipeigwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M.D., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Gibbs R.A., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003544; AAF50040.2; -  
DR FlyBase; FBgn0036181; CG18331.  
SQ SEQUENCE 2586 AA; 260194 MW; 8BBB2435A9FEAE5B CRC64;

Query Match 5.0%; Score 588; DB 5; Length 2586;  
Best Local Similarity 19.0%; Pred. No. 1.9e-16;  
Matches 513; Conservative 369; Mismatches 1085; Indels 732; Gaps 105;

QY 19 VAECASKSHSGSS-----SSTAGOVGSS--PVILRTVATLALIVIGATLNGS 64  
Db 168 VVEVSQGTNGNSSTOSSSTTTTSSDEGQTTSSSDPVVEVAQ-----GSSSNGD 218  
QY 65 AVAQNNSKTAFTGTGNNDNASNEASIAIGSLAKAHANQAIAGSKPDPNROANOKA 124  
Db 219 GNSTQSLTTTTTTTTTSSDGGQSTSSDPVVEVSQ-----GTNGSNSSTQSSSTTTT 271  
QY 125 GSHAKGKESIAIGDV-LAEGDASIAIGSDDLYLDNRNSTNSKYPNGLLSTLIQHTVLRQ 183

Db	272	TSSDEGQTTSSDPVVEAOGSSNG-----DGNSTQS-----STTTTTT---	313
Qy	184	IRDSNGOKYRTAEGCHASTAVCAMAYAKGHFANAFGTRSTAEGNYSLAVGLTAKAERG	243
Db	314	-----TSSDGGQSTLSDPVVEVSQGTNG-GNSSTQS-----SSSTTTTTSSDEG	357
Qy	244	YTAIGSNAQAINYGALGADTRVDLDYGIALGYGSQLNNNNNNNNKAYVPEGNGSNI	303
Db	358	QTT-----SSSDPV--GEVAOGSSN-----GDGNSTQSSTTTTTTTTTSSDGGQSTT	403
Qy	304	KS-----SKATGNGLFSIGSSTIKRKIIYAGYEDTDVAVNQAOLKAVENLAKRQITFK	357
Db	404	SSDPVVEVSQGTNGGNSSTQSSTTTTTSSDGGQSTSSDPVSEVAGGSSI-----456	
Qy	358	GDGNGTGVKKLGETLTIK--GGETQAOKLTDNNNIGVYVTONNTGLVKLAKNLSGLETV	415
Db	457	GDGNSSTQSSTTTTTTTTTSSDGGQS-----TSSDPVVEVSQGTNG-----GNSSTQSSS	506
Qy	416	STNLTASEK-----VTV-----GSGNNTAELOSGGLTFTPTTNA-----450	
Db	507	STTTTTSSDGGQSTSSDPVVEAOGSSNGDGNST--QSSTTTTTTTTTSSDGGQSTT	563
Qy	451	STDKTVYGTDLKFTDNSNTALEDTTRTKIGLFSNKRAGTVDENKPYLD-----KDKLKV	506
Db	564	SSDPVVEASQGTNGGNSSTQSSTTTTTSSDGGQSTSSDPVSEVAGSSSGD	618
Qy	507	GNSLTNNGGLTVNNTIGGSKNOIQVGADGKIEADVNVVNSA---AKGTTTRITEEIGCF	563
Db	619	GNSSTQSSTTTTTTTTTSSDGGQSTSSDPVVEASQGTNGGNSSTQSSTTTTTSSD	675
Qy	564	ADAGKVDKSPYLDKQLOVGVKIKDGINAGDOKISNVKDATDDTDAVYKQLKQV	623
Db	676	---EQGTTSSDPV-----VEVAGSSN-GDG--NSTQSSTTTTTTT-----713	
Qy	624	QODADGALQSFIRDE--KGOEFTISNLYNGNTPNFTTITFAGENGISIN-----DI	676
Db	714	-TSSDGG-QSTLSDPVVEVSQGTNGGNSSTQSSTSSST--TTTTSSDGGTSSSDPVVEV	769
Qy	677	AKGKVKVGI-DP INGLTTPKLTUVSGDKDGKQTLVIEQVAGSNDRKNIIRGLSPLPSITN	735
Db	770	AGGSSNGDGNSTQSSTTTTTTTTTSSDGG-----ESTSSDPVVEVSQGTNGDNSSTQS	824
Qy	736	AGGVRT--EONGTITSDOKSKAA---SIGDILNTGFKLNKNSNSGVFVSYNTVDF	788
Db	825	SSSTTTTTSSDGGQSTSSDPVSEVAGSSSTGD-----GNSSTQSSTTTTTTTDI	875
Qy	789	I-----DGNATTAKVYDETNOTSKVYDVNVDEKTIETLGDNGTKNIGVKVITTL-----	839
Db	876	FRCPQRQRTTTTTSSDGGESTLSDPVVEVSQGT--NGGNSSTQSSTTTTTSSDGG	932
Qy	840	-----TTTNANGKATNFSPTDNDALVNAKDIAENLNTLAKEIHTKGT--882	
Db	933	QTTSSSDPVVEAOGSSNGDGNSTQSSTTTTTTTTTSSDGGESTSSDPVVEVSQGTNG	992
Qy	883	ADTALQTFKVKDQADTDEFITVGKDGCTGKTVNTLKLKGENGLTVATNKDGTVEGIN	942
Db	993	DNSSTQSSTTTTTSSDDE---QOATSSAPVVDLSQSSNGDGNSTQSSTTTTTTT	1048
Qy	943	TQSLKAGDSTTLNKGLSIKNPASNEQIQVGADGVKFAKVDKNGSST--GIDGTSRTIK	1000
Db	1049	TSSD---GDQSTSSDPV-----VEVSQGTNG-----GNSSTQSSTTTTTSS	1089
Qy	1001	DQIGFTGANGS-----LDTTKPHLKDKLKVGEVETNT	1034
Db	1090	DEGOTTSSFGSCRRSHRILFQCPRQOQORRHLPWVAMPPLHT-----PVVEVSQGTN	1145
Qy	1035	GINA---GGKKTNIQSGDITQNSDAVGTGRVYDLKTELESKINSAAKTAQNSLHEFSV	1091
Db	1146	GNSSTQSSTTTTTSSDGGQSTSSAPVSSSTTTTTTTTTSSDGGQSTSSDPVVEVS-	1204
Qy	1092	ADEQGNHFTVSNPYSSTYTSKTSVDITFAGENGITTKVNGVVRVGDQTKGLTPEKLTIV	1151

```
QY 1986 NYLTNPAAIDRINEGIRFFHVNDGNQBPVVQGRNGIDSSASGKHSVAIGFQAKADGE 2045
Db 2251 -----VNGKSSKSSSTTTTSTTKGNKLSLSL-----PEVD 2282
QY 2046 AAVAIGRQTAGNSAIGNAQTGDOSTAI-----GTGNVVTGKHSAGID 2093
Db 2283 ASIAVN-----GDAKASAIKDTNLSKIDLSPLKDALNUNGKSSKSST 2330
QY 2094 PSTVKADNSYVGNNOFIDATQDVFVG-----NNITVESVALGNSAI 2142
Db 2331 TTTTSTNG---SKSKILTPVKVDA-GISIDGGISGSTTKIKITSKNSAVPKASSSF 2386
QY 2143 SAGTHA-----GTQAKKSDGTAGTTTAGATGTVKGFAGQTAGVAVSGASGAERR 2193
Db 2387 KTTTSTSTSSVPKTESYSSSSSKTSNPIRLTLPNINAGISVGGGDSGGSWSKLI 2446
QY 2194 IONVAAGEVSATSDVAVGSOLYKATQGTANATNEIDHRI--HONENKANAGISSAMAMA 2251
Db 2447 KRSTSDETNASDGLSUGSIDSRVTPVPRDLRDIRLGRQSGNDASQNSN----- 2501
QY 2252 SMPQAVIPGRSMVTGGIATHNG-----QGAVAVGLSKLS-----DNGQWV 2291
Db 2502 SWTQATQGSSESLANGAITANGLSLESGSGGVATTIPGSGVGVTCQYPPYWGNGRWV 2560

RESULT 10
Q9RL69 ID Q9RL69 PRELIMINARY; PRT; 2478 AA.
AC Q9RL69;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
DE MRP PROTEIN.
GN MRP.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=97302526; PubMed=9158773;
RA Wu S., de Lencastre H., Sall A., Tomasz A.;
RT "A phosphoglucotransferase-like gene essential for the optimal expression
RT of methicillin resistance in Staphylococcus aureus: molecular cloning
RT and DNA sequencing.";
RL Microb. Drug Resist. 2:277-286(1996).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=97431478; PubMed=9286983;
RA Jolly L., Wu S.W., Van Heijenoort J., de Lencastre H.,
RA Mengin-Lecreulx D., Tomasz A.;
RT "The femR315 gene from Staphylococcus aureus, the interruption of
RT which results in reduced methicillin resistance, encodes a
RT phosphoglucosamine mutase.";
RL J. Bacteriol. 179:5321-5325(1997).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=99265121; PubMed=10332717;
RA Wu S., de Lencastre H.;
RT "Mrp-a new auxiliary gene essential for optimal expression of
RT methicillin resistance in Staphylococcus aureus.";
RL Microb. Drug Resist. 5:9-18(1999).
DR EMBL: Y09927; CAB55329.1.
SQ SEQUENCE 2478 AA; 263029 MW; 6B9859A0D2D023C74 CRC64;
```

Query Match 4.9%; Score 578; DB 2; Length 2478;  
Best Local Similarity 20.8%; Pred. No. 4.7e-16;  
Matches 545; Conservative 329; Mismatches 1033; Indels 708; Gaps 120;

```
QY 67 AQNSKIAFTTGNNDNASASNEASTAIGSLAKAHANQAIAIGSKPDPNRQA-----NQ 122
Db 100 AOPTQPA-----QGNQADPNNAQAQPGNOA-TPANQA-GOGNNOATPNNATPANQOTQ 153
QY 123 KASHAKGKESIAIGGDVLAEE-----GDASI-----AIGSDDLYLDRNSTSKY 166
Db 154 PANAPAAQAAPVAANAQTQDPNASNTGEGSINTTLTFDDPAISTDENQDPTVTVTDK 213
QY 167 PNGILSTLIQ-----HTVLRO--IRDSNGSQYRRTAAEGHASTAVGAMAYAKGHFAN 218
Db 214 VNGY--SLINNGKIGFVNSSELRSMDKKNPNQYQ-----AKGNA- 253
QY 219 AFG-----TRTAGNYS-----LAVGLT---AKAEKGYTIAIGSNAQAINYG 258
Db 254 ALGRVANADSTDHGNGFNGISKTIVNKPDSLIINFMTQTNKQCATNLVKAOKKNTTEL 313
QY 259 ALALGADT-----RVLDLY--GIALGYGSOILNNNNNNNNKAYPE---GN 299
Db 314 ATVNAKTGTAHLKVPFTDADRLDLQIPONTAVADASRITTNKDKYKYYSFIDNVLGS 373
QY 300 GSNIT-----KSSKATGNGLFSIGSSTIKRKIINVAG-----YEDTDAVNVQOLKAV 346
Db 374 GSHLYKNRDLAPAKTNKKEYTINTEIGNNG--NFGASLKADQFKYEVTLPQGVY---V 428
QY 347 ENLAKRQITF-KGDDNGTGKVKKLGETLTITKGGETOQADKLTDNNNIGVVTDNNT- 399
Db 429 NN--SLTTFPNGNEIDSTVLK-----NMTV-NYDONANKVFTTSQ-GVTTARGHTTKEVL 479
QY 400 ----GLKVKLAKNLGLETYSTKNLTASEKVTVSGNGNTAELQSGGLTFTPTTNASTDKT 455
Db 480 FPKSLKLSYKVNANIDT--PKNIDFNEKLTARTASDVV-----INNAOPEVT 526
QY 456 VYGTDLKFTDNSNTALEDTTRITTKDKIGFSNAGTVDENKPY-----LDKDKLKVGNS 510
Db 527 L-----TADPFSVAVEMNKDALQQVNSQVDSNHYTTASIAEYNKLKQADT 573
QY 511 LNNGGLTVNNTIGSNKQIQVGADG-----IKFADVNVNVNSNAAKFQTR 555
Db 574 ILNEDA---NHVKTANRASQADIDGLVTKLQAALIDNOAAIAELDTKAQEKVTAQOQSKK 630
QY 556 ITEEEIGFADADGKVDKSPYLD-KKQLOVGGVKITKDSGINAGDQKISNVKDATDFTA 614
Db 631 VTQDEAALVTKINNDKNNAIKQTTAGVYTTKDNGL-----AVLEQDV 678
QY 615 VTYKQLKQVQDADGAL-----QSFSIRDEKGOEFTISNLYSNGNTPNTFTTFA 665
Db 679 ITPTVRPOAKQDIQAVTTTRKQIKKSNASLODEK-----DVAN-----DKIGKTIKA 729
QY 666 GENGISINDIAKGVKVGIDPIN---GLTTPKLTGSKDGKTQULVIEQVAGSNDTKN- 721
Db 730 DIDAAATTNAQVEAIKTK-AINDINQTTTATTAKAAALEEFDEVVQAQIDQAPLNPDTNE 788
QY 722 ----IIRGLSPTLPSITNAGVTRTEQNTITSDKSKAASIGDILNTGFLNKNNSV 777
Db 789 EVAEAIERNAAKVSGVKAIEATTADLERVNEEKIENITDSTQTKMDA---YNEV 845
QY 778 GFVSTYNTVDFIDGNATTAKVYDETNOTSKVT-----YDVNVDEKTIETLTDGNGKT 829
Db 846 KOATARKAQ-----NATVSNATNEEVAEADAANDAAQKQGLHDIQVVKSQEVAADTKSV 901
QY 830 -NKI-GVKTTTTLTNTNANGKATNFSTTDNDALVNAKDIAENLNTLAKEIHTTKGTADTAL 887
Db 902 LDKINAQIQTAKVKPAADTEVENAYNTRKQEQNS-----NAST-TEEKQAAAYTELDTKK 955
QY 888 QTPKVKKGATDDETTITVGKDGQTQNGKTVTVTLKLGKENGTLVATNKDGTGTFGINTOSGL 947
Db 956 QEARTNLDAANTNSDVTAKDNS-----IAAINVOQAATTK-----KSDA 995
QY 948 KA--GDSTTLNKDGLSLKPNPASNEQIOVGADGVKFAV-----DKGNSSTGI 992
Db 996 KAEIAQKASERKTAIEAMNDSTTEEQQAQKVDQAVVTANADIDNAAANDVDNAKTTN 1055
```



Db 254 ALGRVNDSTDHGNGFNGISKTVNVPDSELIINFMTQMNSKOGATNVLKDAKNTEL 313  
QY 259 ALALGADT-----RVDLDY---GIALGVGSQILNNNNNNNNKAYYPE---GN 299  
Db 314 ATYVNAKTGTAHLFKVPTDADRLDLOFPDNTAVADASRLTTNKDGYKYSFIDNVLFS 373  
QY 300 GSN-----KSKATGNGLFSGSSTIKRKIINVAG-----YEDTAVNVAQLKAV 346  
Db 374 GSHLYVKNRDLAPAKNNKTYTINTEIGNNG--NFGASIKADQFKYEVLTPQGVTV---V 428  
QY 347 ENLAKROITE-KGDDNGTVGVKKLGETLTIKGGETAQDKLTDNNNGVVTNNNT----- 399  
Db 429 NN--SLTTTFPNNGEDSVLK-----NMV- NYDQANKVTFYSO-GVTTARGTHKEVL 479  
QY 400 ----GLVKLAKLNSLETYSTKLTASEKVTVSGSNTAELQSGGLTFPTTNNASTDKT 455  
Db 480 FPKSLKLSYKVVANIDT--PRNIDFNEKLTyrTASDVV-----INNAQPEVT 526  
QY 456 VYGTDLGLFTDNSNTALEDTTRITKDKIGFSNKAQTVDENKPY-----LDKDLKLVGNST 510  
Db 527 L-----TADPFVSVAEMNKNLALQOOQVNSQVDNSHYTTASIAEYNKLKQOQADT 573  
QY 511 LNNGLTVNNTIGSNKQIOVGADG-----IKFADYVNVVNSNAKGTTR 555  
Db 574 ILNEDA---NHVKTANRASQADIDGLVTKLOAALIDNOAAIAELDTKAQEKVTAAGOSKK 630  
QY 556 ITEEEIGFADADGVKQPSYLD-KKQLQVGVKIKTKDSGINAGDQKI-----SNVKDATD 610  
Db 631 VTODEVAALVTKINNDKNAIAEINKOTTAQGVYTEKNGIAVLEQDVITPTVKPRANQD 690  
QY 611 DTAVTYKQLKVOQDADGALQSFSDIRDEKQBEFTISNLSYNGNTPMFTTITFAGNGI 670  
Db 691 IOQAVTRK-QQIKK-----SNASLQDEK-----DVAN-----DKICKIETKAIDIAA 734  
QY 671 SISNDIAKGVKVGIDPIN--GLTTPKLVGSDKQKGTOLVIEQVASGNDTKN-----I 722  
Db 735 TTRAQVBAIKR-AINDINGTTPATTAKAAALEEFDEVQAOIQOAPLNDPTTNEEVAEA 793  
QY 723 IRLSPLTLPISITNAGGVRTTEQONTITSDEKSKAASIGDILNTGFLKNNSSVSGFVST 782  
Db 794 IERINAAKVSQVKAIEATTTAQLDERVKNEEISKIENITDSTQTKMDA---YNEVKQOAT 850  
QY 783 YNTVDFIDGNATTAKYTYDETNOTSKVT-----YDVNVDEKTELTGDNGKT-NKI- 832  
Db 851 ARKAQ-----NATASNATNEEVAEADAADAAQGLHDIOQVVKSKQEVADTKSKVLDKIN 906  
QY 833 GVKTTTTLTNNAGKATNFSTTNDALVNAKDIENLNTLAKETHTTKGTADTALQTFKV 892  
Db 907 AIQOQAKVPAADTEVENAYTRKQEIQNS-----NAST--TEEKQAAYTELDTKKQOART 960  
QY 893 KKDQATDDETITVCKDGTONGKTVNTLKLKGENGLTVATNKGDTVTFGINTQSGLKA--G 950  
Db 961 NLDAANTNSDVTAKDNS-----IAAINQVQAATTK-----KSDAKAIEA 1000  
QY 951 DSTLTKDGLSIRNPASNEQIOVGADGVKFAV-----DKGNSSTGIDGT-S 996  
Db 1001 QKASERKTATEAMNDSTPEEQAAKDKVDQAVVTANADIDNAANNDVDNAKTTNEATIA 1060  
QY 997 RITKD-----QIGFTGANGSLDITTKPHLTKDKLKVGEVEI----- 1031  
Db 1061 AITPDANVKAQAIAADKVOAQETADGNNGS--TTEKAAAKQOQVOTEKTTADAIDA 1119  
QY 1032 -TNTGINAGGK----KITNIQSGDITO-NSNDAVTGGVRYDLKLETSKINSAKTAQNS 1085  
Db 1120 HTNAEVEAAKAAIAKIEAIOPATTTKDNKAIEA-----TKANERKTAIAQODIT 1171  
QY 1086 LHESFVAD-----EQNHFTVSNPYSSYDTSKSDVITTFAGENCI---TTKVNKGVV 1134  
Db 1172 AEETAANADVDNAVTOANSINFAANSQNDVDQAKTT-----GENSIDQVTPTVNK--- 1222  
QY 1135 RVGIDQTKGLTTPKLVGNNGKGIVIDSKDGONTITGLSNTLANVNDGAGHALSOGLA 1194

Db 1223 -----KATARNEITAILNNKLOEIOATPDATDEEKQAADEAEANTENGRKANQAI SAATT 1275  
QY 1195 N-DTDKTPRAASIGDVLNAGFNLOCGNEAVDFVSTYDTVDFIDGNATTAKVTYDDTSKTS- 1252  
Db 1276 NAQVDEAKA-----NAEAAI-----NAVTPKVVKQAADKEI 1307  
QY 1253 ---KVVYDVNVNDKNTIEVTSDDKLGKGVKTTTLTKTSANGNATKFSAAADGDLVLKASDIATH 1309  
Db 1308 DOLQATQNTYINNDONATTEKEAAIOOLATAVATVDKANNIT--AATDDNGVDOAKDAGKN 1365  
QY 1310 LNTLAGDIOIOTAKGASQASSASYVDAD-----GNKVLYDST-----DK 1347  
Db 1366 -----SIQSTQFATAVKSNAKN-DVDQAVTQOALDNTTGATTEEKNAAKDLVLKAKE 1418  
QY 1348 KYQY-----VNDKGQVDKN-----KEVAKDKLVAQAQTPDGTILAOHNVK 1386  
Db 1419 KAYQDILLNAQTNDVTOIKDQAVADIQITADTTIKDVAKDELATRANEQKALIAQTADA 1478  
QY 1387 SVINKEQVN--DANKKQG-INEDNAFIKLENAKATKTKNAAVTVGDDLNAV-AQTPLT 1441  
Db 1479 TTEKEQAOQOVDAHLTQGNQNIENASOISDDVNTAKD-----NAIOAIDPIQ 1525  
QY 1442 FAGDTGTTAKKLGETLT-----IKGGQDTDNKLTDNNIGVV--AGTQGTVKLAKDL 1491  
Db 1526 ASTDVKTNAR--AELLFEMQNKITEILNNETTNEEKGNDIGPVRAAYEEG----- 1574  
QY 1492 TNLNSVNAQGTTRIDKIGSFWDANGQAQKANTPVLSANGL--DLGKKRISNIGA----AVD 1545  
Db 1575 --LNNINAATTCG-----VTTAKDTAVQVQOOLHANPVKKPAGKKELDQAAADKKTOIE 1627  
QY 1546 DNDVNFQKQNEVAKTVNNLNNSGASLFFVYVDANGKPIGTGKPKQAIKAGDGY 1605  
Db 1628 QTPNASOOEINDAKOEVDTELNOAKTNVQO-----SSTNEYVDNAVK--EGK- 1672  
QY 1606 YHANANGVPVDKQKPTDADKLANLAHKGKPLDAGHOVVASLGGNSDATTLTNIKSTLP 1665  
Db 1673 --AKINAVKTFSEYK-----KDALAK-----IEDATNAKVNEADNASTSSEI----- 1714  
QY 1666 QIDTPNTGNANAGQAQSLPSLSAAQQAQSVKIDLVNNGFNLOTHNHQVDFVKAYDTVNF 1725  
Db 1715 -----AEAKQKLAELQATADQNVNOATSKDDIEVOI-----HNDLDNINDY-TIPT 1759  
QY 1726 VNGTGADITSVRSGADGTMSNITVNTALAAATDDGNGVLIKAKDGKFFYKADDLMPNG----- 1780  
Db 1760 GKRESATDLYAYADOKKNNISADT--NATODEKQQAIKOVQDQNVQTALESINNNGVDNGD 1817  
QY 1781 ---SLKAGKSASDAKPTGLSLVNPNAKG-----STGDAV-----ALNNL 1818  
Db 1818 VDDALTOGKAIDAIAQVD--ATVKPKANQAEVKAEDTKESIDQSDQITAEKTEALAMI 1875  
QY 1819 SKAVFKSKDGT--TTTIVSSDGSIOGK---DNSSI--TLSKDGILNNGKVISNVGKGT 1871  
Db 1876 KQITDQAKQGITDATTTAEEVEKAKAQGLEAFDNIQIDSTEKQKAIEELETTALDQIEAGVN 1935  
QY 1872 DTDAAVYQQLNEVRNLLG--LGNAGNDNADGNQVNIADIKKDPNSGSSNRTVTKAGTVL 1929  
Db 1936 VNADATTTEKEAEFTNALEDILSKATEDISD--QTTNAEIAATVKNSALEQ-----LKAORI- 1988  
QY 1930 GKGNNDETEKLGVOVGVDKD---GNANGDLSNVVVKTKDQSKKALLATYNAAGOT 1985  
Db 1989 ----NPEVKNALEIREVNVNKQIEIKKNADADASAKIEARTDLGRYDFDFADKLDKTQT 2044  
QY 1986 N-----YLTNNPABAIDRINBQGRFFHFVNDGNQBPVQVQNGIDSSAGKHSVAIGFOAK 2041  
Db 2045 NAEVAELQNTVIPAIEAIVPQNDP--DANDTN-----NGID----- 2078  
QY 2042 ADGEAAVAIGRQTAGNQSTAGDNOQA-----TGDOQIAIGTGN-----VVTGKHSGA 2090  
Db 2079 -----NNDATANSANATPENTQPNVSETTANGKADASPTTPNNSDAA 2122  
QY 2091 IGPDPSTVKAD--NSYSVGNNOQFIDA-----TQTDVFGVGNNT---VTESN 2132  
Db 2123 TGETTATSATDDANDKFPQANNSSVDASTNSPTMDNDVTSKPEVESTNNGTTDKPVTETD 2182

Thu Sep 13 14:18:21 2001

Qy	399	TGLKVKLAKNLISGLTETVSTKNLTASEKVTGVS--GNNTAELOS-----	439
Db	286	TILK---ATNINLKHNTSTNLND--DNIIIVGNIKGNKKDILLNFKVHGHTNLNDNEMIIPAP	341
Qy	440	---GGLTFTPTTNASTDKTVYGTGDLKFTDNTSNTALEDTTRITKDKIGFSNK---AGTV	492
Db	342	OKTHGTLEF--KGNATLNGNINNLILKFSGGHGTLMLOGNTKYDNLVFPADSVLDSGTI	399
Qy	493	DENKPYLDKDLKLVGNSSTLNNGLTVN--NTI-----GGSNKQIQVAGDGIKFAADVNVV	545
Db	400	SVN-GLLTDCTVTPNNNSVNGGTLIIINAKNTISAKLLNATKAKIQIN-----ANLTMNH	452
Qy	546	SNAAKFGTTRITEEIEIGFADA-DGKVDKKSPLYDKKQLQVGGVKTITKDSG-----INAG	598
Db	453	PSAGDISIRIADNTIYTIIDAKNGVN-----LLNNNAKILFEGADSMALLINTG	502
Qy	599	DOKISNVKDATDDTDAVTKQLKOVQDADGALQSFSTRDEKGOEFTISNLYSNGNTPT	658
Db	503	-----VTADRTFTIYNLNQSGNDEYGVKTEAIK-----	532
Qy	659	FETITFAGENG-ISISND-----IAKGVKVGIDIPINGLITPKLTVGSKDKGTQLV	709
Db	533	-KVITIANQSGPYTIGQDNTHRLKELIVEGAGDIIID--DTIFTKLLSINSTGQITFNT	589
Qy	710	IEQVAGSNDTKN-----IIRGLSPTLPSITNAGGVRTTEQGN---TITSEDKSKAASIG	761
Db	590	LDLGAGGNIAFGKHGTVLVVNGVGTGSIITSENNOGILTINSNIGTIGVIGNELGLKLVNIG	649
Qy	762	-DILNTGFNLKNNNSVGFVSTYNTVDFIDGNATTAKVTYDETNOTSKVTVYDVNVDEKTI	820
Db	650	ADPYTCSANV---FASVALTNPSVLLADGVTLTGEVT---THNNTKGVLSLG-----	697
Qy	821	ELTGDNGKTNKIGVKTTLTTTNNANGKATNES-----TTDNDALVNAK	863
Db	698	--TGSN-ITGOIGTNSAALEKINIGAGASNIDSNIVAGSTVLTDQOTSELTANDVNVNSN	754
Qy	864	DIAENLNLAKETHTTKGTADTALQTFKVKKDGATDDETITVKGDKGTQN-GKTVNTLKLK	922
Db	755	IITTAGNNSKLIIFTGNGGI-----TGNIGANGAALQEVVF---NGTTNIGGTANS-----	802
Qy	923	GENGLTVATNKDGVTFGINTQSLKAGDSTTLNKDG-----LSIKNPASNEOIQVAGDG	977
Db	803	--QNFTVAHSAANVITGLTT-GALKYKDTGTIIAHGVLGVDIDFNKAG-----	849
Qy	978	VKFAKVDKGNSTGIDGTSRITKDOIIGFTGANGSLDTTKPHLTKDKLKVGEVEIT-NTGI	1036
Db	850	-KFILGD---CAMIDGSVLCN-----GGVAGTLDF-----IGDNVTQNLGA	887
Qy	1037	-NAGGKKTINIOGSDITON---SNDN-----VTGGRVYDLKTELESKINSAAKTAQNSL	1086
Db	888	DNANSISTINIO-GDMTKNVTIANDIFVDNIHFTNGGIIQL-----	927
Qy	1087	HEFSVADEQGNHFTVSNPYSYDTSKTSKVITTFAGENGITTKVNVKVVVRVGDQTKGLTT	1146
Db	928	-----GGNLTJHN-----IDF-GANGGTLFEN-----	948
Qy	1147	PKLTVGN--NKGIVIDSKDGONTITGLSNTLANVTNDGAGHALSQGLANDTDTKT-RAA	1203
Db	949	-----GNNTYNLNAIIV---NQN---GILNAFTNLK-----ASDDTIGTVKII	986
Qy	1204	SIGDV-LNAGFNIGQNGEAVDFVSYVD-TVDFIDGNA-TTAKVTVDDTSKTSK-----	1253
Db	987	NIGQIGTPQNFITQVNNKNTLVSSVNSSTNFCDANSQILLSAPVDQTIKFINNLNETGG	1046
Qy	1254	--VYVDVNDKTIETVSDKKLGKVTTLTKTSANG-----NATKFSAA	1295
Db	1047	GIITLDSNGNLTISGNGIKLGSKNELSSLNKGKVTVNDLDIQIHOLNINNGALF	1106
Qy	1296	DGDALVKASDIATHLNTLAG-----DIOTA-----KGASQAS--	1327
Db	1107	DDQSITSAKIKNINIGTVAGGATVTLDAINDNFDLNTSGMVFVKHODSILELKNSSNTNDH	1166
Qy	1328	--SSASYVDADGKN--VIYDSTDKKYYQVNDKGVQDNKKEVAKD--KLVAQAQOTPDGTLA	1381

QY

2133

SVALGNSAISAGTHAGTQAKKSDGTAGTTTATAGATGTVKGFGAGQTAVGAVSVGASGAER

2192

Db

2183

NATPAESTTNNST---TTATNENAPTGSTATAPTASTEAASSADSKONASVNDKQNA

2239

QY

2193

RIQNVAAAGE-----VSATSTDAVNGSOLYKATOGIANA-----

2225

Db

2240

EVNNSAESQSTNDKYAOPKSENKAKAEKGDSDSTNQSWESTETLPSADITPEPNVPSMT

2299

QY

2226

-----TNELDHRIHQENENKANAGISSAMAMAPQAYIPGRSMVTGGIATHNGQAV

2277

Db

2300

SKDKEESTNOTD---AGQLKSETNVASNEADKSPKADTEVSNKPDTSASSEAKEKMT

2355

QY

2278

AVGLSKLSDNGOWYFKINGSADTQGHVCAA

2307

Db

2356

STNLSQKDDTA-----TADTNDTQPSVGSA

2380

RESULT

12

Q92D91

PRELIMINARY;

PRT: 2340 AA.

AC

Q92D91

DT

01-MAY-1999 (TReMBLrel. 10, Created)

DT

01-MAY-1999 (TReMBLrel. 10, Last sequence update)

DT

01-MAY-2000 (TReMBLrel. 13, Last annotation update)

DE

CELL SURFACE ANTIGEN (SCA3).

GN

RP451.

OS

Rickettsia prowazekii.

OC

Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC

Rickettsiaceae; Rickettsiae; Rickettsia.

OX

NCBI\_TaxID=782;

RN

[1]

RP

SEQUENCE FROM N.A.

RC

STRAIN=MADRID E;

RX

MEDLINE=99039499; PubMed=9823893;

RA

Andersson S.G.E., Zomorodipour A., Andersson J.O.,

RA

Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,

RA

Eriksson A.-S., Winkler H.H., Kurland C.G.;

RT

"The genome sequence of Rickettsia prowazekii and the origin of

RT

mitochondria."

RL

Nature 396:133-140(1998).

RL

EMBL: AJ235271; CAA14908.1;

SQ

SEQUENCE 2340 AA; 247862 MW; CC4070F93C165F26 CRC64;

Query Match

4.9%;

Score 571.5;

DB 2;

Length 2340;

Best Local Similarity

21.7%;

Pred. No. 8.2e-16;

Matches 554;

Conservative 295;

Mismatches 935;

Indels 773;

Gaps 133;

QY

59

ATLNGSAYAONNSKIAFGT-TGNNDNASASNEASTAIGSLAKAHANAIAIGGSKPDPRN

117

Db

14

ASLUGTLFTNSN---ATGTLIPNGSVSLNTDAGLVGVFNNGDIQ-IVNGGREIKISA

69

QY

118

QAANOKAGSHAKGESIATGGVLABGDASTAIG---SDDLDRNSTNSKYPNGLLST

173

Db

70

DKANAIIIGGINTLKLPDFGG---VEVSNVSIGPLNAGEDL-----NTNF--GPLKF

117

QY

174

LIONHTVLRQIRDSNGSQYRRRTAEGHASTAVGAMAYAKGHAFANAFTRSTAGBNYSIA

233

Db

118

ISNVTSI-----ITGVGKTFESNIDFA---GKNATLQINKDL-

152

QY

234

VGLTAKAEKGYTIAIGSNAQAINYGALALGADTRVDLDYGA---LGYGSQILN-NNNN

289

Db

153

-NITTKID-----NTVAGNNGSITPEGS-----GILSNHIGYTNLSLLGINVNG

195

QY

290

NKAYVPEGNSNKKSKATGNGLFSIGSSTIKRKIINVAGYEDPDVAVNAQLKAVENL

349

Db

196

EAKIYAPEANNITNAKN-----

225

QY

350

AKRQI-TFKGDDNGT-----GVKKKILGETLTIKGETQADKL-TDNNGNIGVVTDNN

398

Db

226

CDGNITTLKGINNTTETDGGIILNLAYDLGSSIIITGIDIGNSLDTINVILGSAFENS

285





Thu Sep 13 14:18:21 2001

```

1847 SSTITLSDGL-----NVGGKVISNVGKTKDQTDAAVQOLNEVRLNLLGLGNAGNDNADG 1900
1281 ANVLFKDAVQLTQTGNIIGGFLDFNAKNGT-----VTLN----- 1314
1901 NQVNIADIKDPNKGSSSSNRRTVIKAGTIVLGGKGNNDTEKLA---TGGVOVGVDKGNAN- 1956
1315 NNVNVAAGTVK--NIGTNGTLLI-----VLGASLNRVNGIAMLKVGAGNVTTIAGGNVKI 1368
1957 GDLSNVVVKTKQDGSKALLATYNAAGQTYLNNPAPAEIDRINPQGIIRFFHVNDGNOEP 2016
1369 GEIQTGTINT-----LTLPAHFKLGT-----SINKTGGALKLNFMMGSSVSG 1411
2017 VYQNRGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTQAGNQSTAIAGNAQTGDSIA 2076
1412 VV-----GTAANSVGDITTAGATSPASSVNAKGTATLGGT 1446
2077 IGTGNVTVKHSIGAIGDPSTVKADNSVSVGNNOQFIDATQTDVFGVGNITVTESNVAL 2136
1447 TSFANTFT--NTGAV-----TLAKGSITSFAKN---VTATSF---VANSATINFGNSLAF 1493
2137 GNSAISAGTHAGTQAKKSDGTAGTTTTAGA-----TGTVK-----GFAGQTAVGA 2182
1494 NSNITGS-----GTTLTGANOVTYTGTSFTDTLTLNTTFDGAASGSG 1537
2183 VSVGASGAERRIQNVA--AGEVSATSTDAVNGSOLYKAT 2219
1538 NILIKSGSTLDLSGVSNLALVVVTAINFDMNISPDTKYT 1576
RESULT 14
Q55582 PRELIMINARY; PRT: 3029 AA.
AC Q55582;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 308.8 KDA PROTEIN.
GN SLR0364.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RC STRAIN=PCC6803;
RA Tabata S.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
[2]
RN [2]
RC STRAIN=PCC6803;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
[3]
RN [3]
RC STRAIN=PCC6803;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D63999; BAA10087.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 3029 AA; 308789 MW; BBE0BB57E822A4FB CRC64;

```

Query Match 4.8%; Score 560; DB 2; Length 3029;  
Best Local Similarity 21.1%; Pred. No. 3.5e-15;  
Matches 562; Conservative 320; Mismatches 985; Indels 794; Gaps 131;

QY 56 VIGATLNGSAYAONN-----SKTAFGTTGNNDNASASNEASIAIGSLAKAHANQ 104  
DB 432 LIDANATGGGYFEQEPFQDTLGNPLISLNOFALTGLDVGNSSTTEREFYEV-----GYSNY 488  
QY 105 AIAIGGSKPDRNOAOKAGS-----HAKGESIAIGDVLAEGDASIAIGSDDLYLDRN 160  
DB 489 QIAVVG---PPRLNVSESTGGDKTLRVDGRPSVLGT-----ISFDD-----527  
QY 161 STNSKYPNGLLSTLIONHVVLRIDRSNGSQYRRPAAEGHASTAVCAMAYAKGHPANAF 220  
DB 528 -TASFIANFL--TPVSSLTVRAGVWNNASNDVRQ-----FSFALGA-----PL 568  
QY 221 GTRSTAGNYSVLAAGLTAKAEKYTTAIGSNAQAIKYGALALGADTRVDLDYGIALGYG- 279  
DB 569 GTFNSPQTFDTIAPIVTSILRNPTSEI--TNADSLTFQVTFSENVQNVVDNDFILSGTV 627  
QY 280 -----SQILNNNN-----NNNKAYVPEGNGSNKSSKATGNGLFSGSSTIKRK 324  
DB 628 SGATISNVSSGGSVYTVTVTGIANNGTVNLDFASGNIRD---TANNALSNTPPTTDEQ 684  
QY 325 I-----INVGAGYEDTDVAVNAQLKAVENLAKRQITFFKGDONGTGVKKKLGTEL 374  
DB 685 YTLDTAPRASITLDANITADDIINIAE-----SQCAIPITGT---VGGEFNVGDTVT 734  
QY 375 I-----KGG-----ETQADKLTDNN-----NIGVVTDNNTGLKV 403  
DB 735 LTVNDKTFGAVGAGGLFSLNPGSLLIVDADULTIAASTATTDAAGNLGSAATNQ-----790  
QY 404 KLAKNLSGLETVSKNLTAASEKV-----TVSGGNN-----TAELOSGGLTFTPT 447  
DB 791 YTVDTTAPIITVNDVADNIIINAAESQAIPITGVGGEFNVGDTVTLTVNGKPTGT 850  
QY 448 TNASTDKT--VYTDGLKFTDNTALEDTTRTKKIG-----FSNKAGTVDENKPYLD 500  
DB 851 VDANGDFSIDLVLGDLVNGSDLTIAASVATT-----DAAAGNPGSASDNQTVTDTTAPT 906  
QY 501 KDLKV-----GNSLNLNGGLTVNNTIGSNKQIQVGADGIKFADVNVN 544  
DB 907 INAIIVDDIINAVEAGSPVAVSGTTTGVGEGQVTVTVTDGNTTATVTVGNAWTF-----NIP 963  
QY 545 VSNAAKFGTTRITEBIG-FADADGK-----VDRKSPYLDKQQLQGVGKITKDS 593  
DB 964 VADIANF---EATEEVVATVSDLAGNPATPATRNIIVDIVAP-----AVTIDSIDDT 1013  
QY 594 GINAGDQKISN-----VKDATDDTDVATYKQKOVQDQDAGALQSFSIRDEKQGEFTISNL 649  
DB 1014 GAQAND-FITNDDTLVFNCTAEADSTVWVSLDGE-----IGTV 1051  
QY 650 YSNGTNPTE---TITFAGENGISINDIAKVKVGDIDPINGLETLTPKLTGVSDKDGKT 706  
DB 1052 TANGAGEWTLDTYGTLLADGYELSVT-----ATNPTGNSATATQTVIVDVTAPT 1101  
QY 707 QLVIEQVAGSNDTKNIIRGLSPLTSPITNAGGVRTTEQGN--TITSDEKSKAASIGDIL 764  
DB 1102 -VTINAIIV-DDIINAVEAGSPVAVSGTTG---VEDGQVTVTDGNTYATVTVGNAW 1155  
QY 765 NTCFNKLNKNSNVGFSTVYTVNDFDGN---AHTAKVYTDENOTSKVYDVNWDE-----817  
DB 1156 TFINPADIANFEATEBEVATVSDLAGNPATPATRNIIVDVTPTVIN-AIAVDDIINA 1214  
QY 818 -----KTIELTG-----DNKTNKIGVKTTLTTLTNANGA-----TNFSTTDNDA 858  
DB 1215 VEAGSPVAVSGTTTGVGEGQVTVTVTDGNTYATV-TVGNAWTFNIPVADIANFEATE-EV 1272  
QY 859 LVNAKDIAEN-----LNTLAKEIHTTKGTADTALQTFKVKKDGADDEITVVGKD 908  
DB 1273 VATVSDLAGNPATPATRNIIVDVAVAVTIDSIDDTGAQ-----NDFITNDDTLVFN-N 1326  
QY 909 GTONGKTVNTLKLKNGENGLTVATNKGDTVTFGINTOSGLKAGDST-----TLNKG-----959

DB 1327 GT-----AEADSTVWVSLDG-IEIGTVTANG--AGEWTLDTYGTLLADGYEL 1371  
QY 960 -LSIKNPASNEQIQVGADGVKFAKVDKNGSSTGIDGTSRITKQDQIGFTGANGSLDTTKPH 1018  
DB 1372 SVTATNPT-----GNSAT---ATQTV-----VDTTAPT 1397  
QY 1019 LTRDKLKVGE-----VEITNTGINAGKKITNIQSGDITQNSDNDAVTCGGVYD 1066  
DB 1398 VTINAIIVDDIINAVEAGSPVAVSGTTTGVGEG---QVVTVTDGNTYATVTVGNAWT 1452  
QY 1067 LKTELESKINSAAKTAQNSLHEFSVADEQNHFTVSNPYSYDTSKTSVDVTFPAGENGIT 1126  
DB 1453 FNIPVADIANFEATEEYVA---TVSFLAGNPATPATRNIIVDVTPTVINAIIVDDII 1508  
QY 1127 TKYNGKVVRVIDQTKGLTTPKLTGVNNGK--GIVTDSKDGQNTIGLSNT-----LANV 1180  
DB 1509 NAVEAG-SPVAVSGT-----TTGVGEGQVTVTVTDGNTYATVTVGNAWTFNIPVADI 1559  
QY 1181 TN-----DCAGHALSQGLANDTDKTRAASIG-DVLNAGFNLQNGEAVDEVST 1227  
DB 1560 ANFEATEEVVATVSDLAGNPATPATRNIIVDVAVAVTIDSIDDTGAQAN---DFITN 1615  
QY 1228 YDTVDFIDGNA---TTAKVTYDDT---SKTSKVYVDVNDV-NKTIEVTSKKLGKVTTL 1280  
DB 1616 DDTLVF-NGTAEADSTVWVSLDGIEIGTVTANGAGETWTLDTYGTLLADGYELSVTATNP 1674  
QY 1281 TKTSANGNATKFSNAODGDLVAKSDIATHLNLTLAGD--IOTAKGASOASSASVYDADGN 1338  
DB 1675 TGSATATQ-----IVDVTAPTPTVINAIIVDDIINAVEAGSPVAVSGTTTGVGEG 1726  
QY 1339 KVIYDSTDKKYYQVNDKQVQDNKKEVAKDLVAAQATPDGTLAQMNYKSVINKBQVNDAN 1398  
DB 1727 QVTVTIDGNTYATVTVG-----NAWTFNIPVADIANFEATEEVV 1766  
QY 1399 KKGINEDNFAIKLENAAKDKTKTNA---AVTVGD-LNAV-AQTPLTFFAGDTGTTAKKL 1453  
DB 1767 ATVSDLAGNPATPATRNIIVDVTPTVINAIIVDDIINAVEAGSPVAVSGT---TTGVED 1824  
QY 1454 GEFLTIK-GGOTDNKLTDN---NIGV-----VAG-----TDGFT 1484  
DB 1825 QQVTVTIDGNTYATVTVGNAWTFNIPVADIANFEATEEVVATVSDLAGNPATPATRNI 1884  
QY 1485 VKLAKDLTNLNSNAGTRIDEKGISFVD-----ANGQAKA-NTPVLSANGLDGG 1534  
DB 1885 VDVAVAVTIDSIS-----DDTGAQANDFITNDDTLVFNCTAEADSTVWVSLDGEIGT 1938  
QY 1535 KRISNIGAAVDDNDANDVNFKOFNEVAKTVNNLNQSNOSGASLPFVYVTDANGKP-----1586  
DB 1939 VTANGAGEWTLDTYGTLLAD-GDYELSVTATNPTGNSATATQTVIVDVTPTVINAIIV 1997  
QY 1587 ---INGTDGKPKAKIGA-----DGKYYHANANG-----VPV-----1615  
DB 1998 DDIINAVEAGSPVAVSGTTTGVGEGQVTVTVTDGNTYATVTVGNAWTFNIPVADIANFEA 2057  
QY 1616 -----DKDGKPIIDA-----DKLANLAHAKP-----1637  
DB 2058 TEEVWATVSDLAGNPATPATRNIIVDVTPTVINAIIVDDIINAVEAGSPVAVSGTTG 2117  
QY 1638 LDAGHQVVASLGNSDAITLTNIKST--LPOIDTPN-----TGNANAOAQ 1681  
DB 2118 VEDGQVTVTVTDGNTYATVTVGNAWTFNIPVADIANFEATEEVVATVSDLAGNPATPATR 2177  
QY 1682 SFLSLSAQAQNSNAASKVDLVNLFNLOTNHOVDVFKAYDTVNFVNGTG-ADITSVSAD 1740  
DB 2178 NITVDTVAVAVTIDSISD-----DTGAQANDFITNDDTLVFN-NGTAEADSTVWVSLD 2228  
QY 1741 G-TWSNITVNTALAAATDDGNNVLIKAKDGKFKYKADLIMPNGSLKAGKASDAKTPTGLSL 1799  
DB 2229 GIEIGTVTANGAGETWTLDTYGTLL--ADGDY-----ELSVTATNPTGNSATATQ-----I 2277  
QY 1800 VNPNAGKSGTGDVAVALNLSKAVFKSK-----DGTFTTFTVSSDGISIOGKNSSITLSKDG 1855

```
Db 2278 VDDTAPTITINAIIVDDIINAVEAGSPVAVSGTTT-----GV-----EDGQVVTVTIDG 2327
Qy 1856 LNVGKVISNCK-GTKDTPDAANVQOLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNS 1914
Db 2328 NTYTATVTGNATWIPVADIANFEATEEV-----VATVSDLAGNPAT 2370
Qy 1915 GSSSRVT-----IKAGTVLGGKGNNDTEKLTATGGVQVG-----VD 1950
Db 2371 PATRNITVDDTAPTITINAIIVDDIINAVEAGSPVAVSG-----TTTGVDEQVVTVT 2423
Qy 1951 KDN-----ANGDSNVWVKTKQDKSKKALLATYNAAQOTNLTN 1990
Db 2424 IDGNTYATVTGNATWIPVADIANFEATEEVVATVSD-----LAGNPATPATRNITV 2477
Qy 1991 N---PAEAIDRINEQ---GIRFFHND-----GNOE---PVVOGRNGID-----SSASG 2030
Db 2478 DPAVAVTIDTSDDTGAQANDFTINDTLVFNFGTAEADSTVVVSLDGIETGTANGAG 2537
Qy 2031 KHSVAIGFOAKADGEAAVAGRQTOAGN-----OSIAIGDNAQATGDQSIAT----- 2077
Db 2538 EWTLDYTGTLADGYELSVTATNPTGNSATATQTIIVDDTAPTITINAIIVDDIINAVE 2597
Qy 2078 -GTGNVWTKHSGA-IGDPSTVKAD-NSYS---VGNKNQF-----IDATQ 2116
Db 2598 AGSPVAVSGTTTGVDEQGVVTVTIDGNTYATVTGNATWIPVADIANFEATEEVVATV 2657
Qy 2117 TDVFG-----VGNNTVTES-----NSVALGNSAISAGTHAGTQAKKSD 2156
Db 2658 SDLAGNPATPATRNITVDDTAPTITINAIIVDDIINAVEAGSPVAVS-GTTTGV-----D 2712
Qy 2157 GTAGTITTAG--ATGTVKGFAGOTAVGAVSGAGGAERRIQNVAAGEVSATSTDAVNSQ 2214
Db 2713 GOVVTVTIDGNTYATVTGNATWIPVADIA-----NFEATEEVVATVSD-LAGNP 2763
Qy 2215 LYKATQGI-----ANATNELD 2230
Db 2764 ATPATRNITVDDTAPTITINAIIVDDIINAVEAGSPVAVS----- 2784

RESULT 15
Q9L800 PRELIMINARY; PRT: 2747 AA.
AC Q9L800;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE RTX PROTEIN.
GN ASX.
OS Aeromonas salmonicida.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
OX NCBI_TaxID=645;
RN [1]
RP SEQUENCE FROM N.A.
RC Strain=ATCC 33658;
RA Braun M., Frey J., Kuhnert P.;
RT "280 kDa RTX protein of Aeromonas.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF218037; AAF27914.1; -
DR InterPro: IPR001343; -
DR InterPro: IPR002035; -
DR InterPro: IPR002048; -
DR Pfam: PF00353; hemolysinCabin; 3.
DR PRINTS: PR00313; CABDNCRPT.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS00330; HEMOLYSIN_CALCIUM; 4.
SQ SEQUENCE 2747 AA; 280202 MW; 208FE380E44A5F37 CRC64;
```

Query Match 4.8%; Score 558.5; DB 2; Length 2747;  
Best Local Similarity 20.3%; Pred. No. 3.6e-15;  
Matches 571; Conservative 350; Mismatches 983; Indels 907; Gaps 140;

```
Qy 7 VIFNKATGTFMAVECAKSHSGSSSTAGQVSSPVIRLTVATLAILVIGATLNGSAY 66
Db 12 VVVTQLKGNLYLVAA-----DGSOKLAEGDI-----LPRDAVL-ITPEGASFNG--- 55
Qy 67 AQNSKIATGTTGNNDNASASNEA-----STAIGSLAKAHANOAIATIGSKSPDRNOA 119
Db 56 --GNQTFLLSPTNEQQAEDTSQLAQNODVATGTPDEISALQQAAILGGADPTQAFEA 113
Qy 120 ANOKAGSHAKGKESI-----AIGGDVLAEGDASI-----AIG 151
Db 114 A-RAGGAPAGGNGNIGCVAGSNGGFVTIDRTGDTTIAAAGFDANOTDAGVTGDAAG 172
Qy 152 SDDLY-----LDNRNSTKYPNGLLST-LIQNTVLRQIRDSNGSKYRRTAEAGH 201
Db 173 EDDPFDLVPPVITVASPDNTDPTLTGTDDAPAGSTVTLVLTDDANGNOQTTLTTVNPD 232
Qy 202 ASTAVGAMAYAKGHFANAFGRSTAEAGNSIAGVLTAKAEKGYTI-----AIGSNAQAIN 257
Db 233 GTFSVDVVT-----PLADGSYTVTANVTPDAGNTATATDGGSDVSTAPSTIV 279
Qy 258 GALALGADTRVLDYDIALGYGSOI-----LNNNNNNNNKAYVPEGNGNIKSS----- 306
Db 280 DAPDNTDPTITGTTDAAGSSTVTLVVTDDANGNOQTLSATVQPNGTYSYDTSPLPDG 339
Qy 307 -----KATNGLESIGSSSTIKRIINVGAGYEDTDVAVNAOLKAVENLA 350
Db 340 DYDVTSATVDPAGNTGTATDGGSDVDVTAPT-----AITLDANITDDDDVINAEE-----A 389
Qy 351 KRQITFKGDDNGTCVKKLGELT-----IKGGETQAD----- 383
Db 390 GQQIPVSGTSG-----EPKAGDTVLTVNGKPTGPDADGRFTILVAGSLLAADRTDID 446
Qy 384 -----KLTDNNNIGVVTDNNNTGLKVLAKNLGLETSTVSTKNLTASEKVTVGS 430
Db 447 ASVTSTDAAGNSATATSEGYGVDTPAPDLAITLDANITDDDDVINAEE--ACQOIPV-S 503
Qy 431 GNNTAELOSGLTETPTN-----ASTDKT-----VYGTDLG 462
Db 504 GTVTGEFKAGD-TVTLTVNGKPTGPDADGRFTILVAGSLLAADRTDIDASTVSTDA- 561
Qy 463 KFTDNTSNTALE-----DTT-----RITKDKIGFSNKA-----GTVDENKPY 498
Db 562 --AGNSATATDSEGYGVDTPAPDLAITLDANITDDDDVINAEEAGQOIPVSGTV----- 613
Qy 499 LDKKLKLVGNSTLNGGLTVNN-----TIGSNKQIQVAGDGIRFADVNV 543
Db 614 --SGEFKAGDTVT-----LTVNGKPTGPDADGRFTILVAGSD-----LAASDRTIDASV 663
Qy 544 NVSNAKFGTTTRITEEIGFADAGKVDKSPYL--DKQLQVGVKITKD----- 592
Db 664 TSTDAAGNSTTKSDTENY-----TVDTVAPDLGIDLDPIAVGCDNVVNAQAEADGKTPV 716
Qy 593 --SGTNAGDOKISNVKDATDDTDAVYKOLQVQDQADGALQ-SFSIRDEK-----GQ 642
Db 717 TILSGTVNGD--ANVGDTVTLTLGDSKLTQVTVLGNLQGLGFSTSTADKLGVSSVTA 773
Qy 643 EFTISNLVSNGTPTNPTFTTFA-----GENGISISNDIAKGVKVGVID 686
Db 774 EISVTDAAAGNSTTKSDTENYTVATTAPALGIDLDPIVVGCDNVVNAQAE--ADGKTPVTL 831
Qy 687 -PINGLT-----TPKLTVGSDDKQTLVIEQVAGSN-----DTKNIINGLSPTLP- 731
Db 832 GTVNGDANVGDTVTLTLG--DG-SKLTQVTVLGNLQGLGFSTSTADKLGVSSVTAET 887
Qy 732 SITNAGGVVTTTEQGN-----TITSDEKSKAAASIGDILNTLGNLKNNSNVG 778
Db 888 SVTDAAGNSTTKSDTENYIVDVAAPAPTITIDAN-----NDQLLTKEI--GSDQIQ 940
Qy 779 FVSTVNTVDFIDGNATTAKTVDENOTSKVTYDVNVDEKTIETL-----GNGKTN 830
Db 941 VRADVNHQDLAAG-----GKVTLTITNGS-----ERTVELTLKADGTLQPSDGNAY 987
Qy 831 KIGVKTTTL---TTTNANGKATNFSTTDNDALVNAKDAENLTLAKEIHTTKGTADTAL 887
```

Db 988 SYNAVGVIGWTTTANGKSLVTVATQTDKAGNVSLPQSDTATVLE-----APTADK-- 1042  
QY 888 QTEKVKADGATDDETTVTKG-----DGTONGKVTNLTCLKGE-----NGLTVATNKO----- 934  
Db 1043 -----QATGOEDAASIVPSLSDIDGTVFTIOSLPNTNGTLFLGSVAVANQISITLA 1097  
QY 935 --GTVFTGINTQSLKAGDST-----TLNKDGLSIKKNPASNEQTOVGADGVKFAVDKGN 988  
Db 1098 DAGSLTFKPPNNW-----GDTSPKYVAMNDGLADPTPAT-----LAIKVESCDN 1143  
QY 989 STGI-----DGTSRITKDOIQTGANGSLDTTKPHLTOKDKLVKGEVEITNTGINA 1038  
Db 1144 APTLTISNDVAVSEGLPSGLKDEL---GEDGQKDATDEVVARGVITVGDVDSNLTVSL 1200  
QY 1039 G---GKKITNIOQSDIT-----QNSNDVATGGRVYDLKTELESKINSAAK 1080  
Db 1201 GWTGADVKSVYSDGVKVEWSWDVSGSKTLIGYAGSKDNV-----LEVKLTPVGS-- 1252  
QY 1081 TAQNSLHEFSVADEQNHFTVSNPYSYDTSKTSVDITFAGENGITTKVKNKVVRVIGIDQ 1140  
Db 1253 -----KGDWGEVTLKQPLDHADTTQ-----EDTIDLKINVSVDGALS 1292  
QY 1141 TKGLTT-----PKLTVG-----NNGKGIVID-----SKD 1165  
Db 1293 TKPLTTTVEEDAPLAGMPDAVSVAKTIDPKVLTLGLFDLTQHSATANQRELDLDFKITAK 1352  
QY 1166 GQNTITGLSNTLANVTNDGAGHALSOG-----LANDTD-----KTRAA 1203  
Db 1353 GFKSATDSALIAADVASTAGLVAGSNPYNHLANEVFERFDNGKSASEELIVKLDPG 1412  
QY 1204 SIGDVLNAGNLNGEANDVFSTYDTPFDIGNATTAKVYDTSK----- 1250  
Db 1413 TVAYGANIKFSQMFGEKESGV-----VEFYRGGLVLISTLTFFSSDAAGGEYQNFOVQOG 1467  
QY 1251 --TSKVYDVNDVNDKTEVTSO---KKLGVTTLTKTSANGNATKFS---AADGDALVKA 1303  
Db 1468 GFTMTWIKATNNGNTNISNDSFTKSIIEPLGTTTPOATAYSGTVAPOMGADGKG--- 1523  
QY 1304 SDIATHLNTLAGO---IQAKGASQASSAS-----YVDADGNKVIY-----DSTDK-K 1348  
Db 1524 -----HIDLLGSNEVLMTATGEISISSKAGLPNTLYGEASDGLVFRLEFTPATGKWE 1378  
QY 1349 YQVND-----KGQVD-KNKEVAKOKLVAQAOTPGTLQAMNVKSVINK--EQVNDANK 1399  
Db 1579 FFOYQNMRLPGDGDIDFQIKVVAD-----GSSLSGFATKPLVTPVQSVSSGNA 1630  
QY 1400 KQINEDNAPIKLENAKDTKNAAVTVGDLN-----AVAQ--TPLTFAGDTGTTA 1450  
Db 1631 EGGNLSHTVT--LSDQTHEATQYFAIOGSGANPASSSDWGAQFSNGVYNSATGKIT 1688  
QY 1451 KKLG-----ETLIK-GGQTDNKLTDNNIGVVGAGTGDFTVKLAKD 1490  
Db 1689 VPAGVSGFTVTLTVNDRLVETTETLTVTVGQNGTGTILDNDRAPTA----- 1736  
QY 1491 LTNLSNVNAGGTIDEKG-----ISFVDANQOA-----KANTPVLSANG-----LDLG 1533  
Db 1737 -----AGCHSAGQEDKQVLTWAQFGASDQAVTDLGVKINT--LPLNGTLEVINA 1786  
QY 1534 GKRIISNIGRAVDNDVAVN-----FKQFNEVATVNN-----LNNQNSGASLPFVVTD- 1581  
Db 1787 GQWVAVSAGALISYDEVNAGHLRFVPLGNLSESSNVNNGGALLGNLKGDYASIGYQISDG 1846  
QY 1582 ---ANGK-----PINGTDGKPOKAIGADGKYHYHANANGVP----- 1614  
Db 1847 ANLSNGSKLVIDVAADVAKTNNVT-----LTNGIPSYSPSPSPASITTELFQSGN 1898  
QY 1615 VKDQKPIITDADKLANIAHKGPL--DAGHQVVASLGGNSDAI-----TLNLIKSTLP 1665  
Db 1899 FNKTAFCITNS--LADSSSQEQVLGNSGNDYIVSTHGGDILLYGOGDNLVGGNAT-- 1954  
QY 1666 QIDTPNTGNAN-----AGQAQSLPSLS-----AAQQNAASVKD-----V 1700

Db 1955 QGDSLYGGDGNVLYSLGSDGLYGGAGTDAVLTGNRADYTIKKKSGSYSVNDRWFDFSV 2014  
QY 1701 LNVGFNLQTNHNOVDFVKAYDTVNFVNGTCADITSVRSADCTMSNITVNTALAAATDDGN 1760  
Db 2015 TEKAGITKALHDVEYVQFADGIYQLNQSGELTLPQ---TVVDYPLELSASLSDRGS 2071  
QY 1761 VLIKAKDGFYKADDLMPNGSLKAGKSASDAKPTTGLSLVNPAGKSGTGDVALNLSK 1820  
Db 2072 EVL-----DSLVLSG-----MPKGAELYSNQLLGTVG----- 2099  
QY 1821 AVFKSKDCTTTTSSDGIISIQKD--NSSITLSKDLNNG-----CKVISNVG-----K 1868  
Db 2100 -----VDGKLTLAGSGLWSASALDKVLSGLTLRVPSSAGQVLDKVEATAKEVGTQDTS 2154  
QY 1869 GTKDTDAANVQQLNEVRNLGLGNAGNDN-----ADGNQVNIADIK 1909  
Db 2155 SASDEDSIRLSYFNATE-----GEPGNQNTFGSEHNIVVGDLDGSLVLPQNYNIAFMV 2209  
QY 1910 KDPNS-GSSSNRTV-----IKAGTVLGGKGN---NDTEKLATGGVQVGVKDKG 1953  
Db 2210 DSSGSIKSDAMKTIKSOLAQOVFASLKSSAGTGDGAGKVNIFLVDFOVKNKSISVDL-KDP 2268  
QY 1954 NANGDLSNVWVKTKDCKSKALLAT-YNAAGQI--NYLTNNPAEADIRNEQGIKRFPHV 2010  
Db 2269 DALNQLQNI-IKTM-DGSEKNGGATXYEDVKTANWFKSDFAKL--NTNAKNVAYF-IT 2323  
QY 2011 DGNQEPVVOGRNGIDSSASCKHSAVIGFOAKADGAAVAIGR---OTOAGNOSIAIGDN 2066  
Db 2324 DGQ-----PNTFNSESNPSIGYQWEGKWSAVYLSQIVTGYEQGEYLLDG-- 2370  
QY 2067 AQATGQDSIAIGTGNV--VTGKHSGAIG-----DPSTVKADNSYSVGNNNQ----- 2110  
Db 2371 -----RLVITTTQVYVDPGRWKGVNRPDKGKYEAVTLSGDRD---GNSDQONARDGYA 2421  
QY 2111 -FIDATQI-DVFGVGNNTVTE-----SNSVALGSNSAISAGTHACTQAKKSDGTAGTT 2162  
Db 2422 LLTAAGVTYDAIGIGDQLSVDDLKGFSDGKVLGTGISADGLANAILGTPVSNLPGTDRLD 2481  
QY 2163 TTAG---ATGTVKGFAGQATAVGAVSVGASGASERRIONVAAAGEVSATS-TDA 2209  
Db 2482 GGAGDDILLFGDAVHFAGINGEGYAAV-----KQYVAGKLSAGSVTDA 2523

Search completed: September 13, 2001, 12:57:20  
Job time: 890 sec

us-09-361-619-11.rspt

Thu Sep 13 14:18:21 2001

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 12:39:38 ; Search time 18125.9 Seconds  
(without alignments)  
3635.974 Million cell updates/sec

Title: us-09-361-619-5  
Perfect score: 6972  
Sequence: 1 ccattgatattggcagggtgt.....ccgtgacctgcaggtcgac 6972

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_est3.\*
- 4: gb\_est4.\*
- 5: gb\_est5.\*
- 6: gb\_est6.\*
- 7: gb\_est7.\*
- 8: gb\_est8.\*
- 9: gb\_est9.\*
- 10: gb\_est10.\*
- 11: gb\_est11.\*
- 12: gb\_est12.\*
- 13: gb\_est13.\*
- 14: gb\_est14.\*
- 15: gb\_est15.\*
- 16: gb\_est16.\*
- 17: gb\_est17.\*
- 18: gb\_est18.\*
- 19: gb\_est19.\*
- 20: gb\_est20.\*
- 21: gb\_est21.\*
- 22: gb\_est22.\*
- 23: gb\_est23.\*
- 24: gb\_est24.\*
- 25: gb\_est25.\*
- 26: gb\_est26.\*
- 27: gb\_est27.\*
- 28: gb\_est28.\*
- 29: gb\_est29.\*
- 30: gb\_est30.\*
- 31: gb\_est31.\*
- 32: gb\_est32.\*
- 33: gb\_est33.\*
- 34: gb\_est34.\*
- 35: gb\_est35.\*
- 36: gb\_est36.\*
- 37: gb\_est37.\*
- 38: gb\_est38.\*
- 39: gb\_est39.\*
- 40: gb\_est40.\*
- 41: gb\_est41.\*
- 42: gb\_est42.\*
- 43: gb\_est43.\*
- 44: gb\_est44.\*
- 45: gb\_est45.\*
- 46: gb\_est46.\*
- 47: gb\_est47.\*

- 44: em\_esthum10.\*
- 45: em\_esthum11.\*
- 46: em\_esthum12.\*
- 47: em\_esthum13.\*
- 48: em\_esthum14.\*
- 49: em\_esthum15.\*
- 50: em\_esthum16.\*
- 51: em\_esthum17.\*
- 52: em\_esthum18.\*
- 53: em\_esthum19.\*
- 54: em\_esthum20.\*
- 55: em\_esthum21.\*
- 56: em\_esthum22.\*
- 57: em\_esthum23.\*
- 58: em\_esthum24.\*
- 59: em\_esthum25.\*
- 60: em\_esthum26.\*
- 61: em\_esthum27.\*
- 62: em\_esthum28.\*
- 63: em\_estin1.\*
- 64: em\_estin2.\*
- 65: em\_estin3.\*
- 66: em\_estin4.\*
- 67: em\_estin5.\*
- 68: em\_estomi.\*
- 69: em\_estom2.\*
- 70: em\_estov1.\*
- 71: em\_estov2.\*
- 72: em\_estpl1.\*
- 73: em\_estpl2.\*
- 74: em\_estpl3.\*
- 75: em\_estpl4.\*
- 76: em\_estpl5.\*
- 77: em\_estpl6.\*
- 78: em\_estpl7.\*
- 79: em\_estpl8.\*
- 80: em\_estpl9.\*
- 81: em\_estpl10.\*
- 82: em\_estro1.\*
- 83: em\_estro2.\*
- 84: em\_estro3.\*
- 85: em\_estro4.\*
- 86: em\_estro5.\*
- 87: em\_estro6.\*
- 88: em\_estro7.\*
- 89: em\_estro8.\*
- 90: em\_estro9.\*
- 91: em\_estro10.\*
- 92: em\_estro11.\*
- 93: em\_estro12.\*
- 94: em\_estro13.\*
- 95: em\_estro14.\*
- 96: em\_estro15.\*
- 97: em\_estro16.\*
- 98: em\_estro17.\*
- 99: em\_estro18.\*
- 100: em\_estro19.\*
- 101: em\_estro20.\*
- 102: gb\_est25.\*
- 103: gb\_est26.\*
- 104: gb\_est27.\*
- 105: gb\_est28.\*
- 106: gb\_est29.\*
- 107: gb\_est30.\*
- 108: gb\_est31.\*
- 109: gb\_est32.\*
- 110: gb\_est41.\*
- 111: gb\_est42.\*
- 112: gb\_est43.\*
- 113: gb\_est44.\*
- 114: gb\_est45.\*
- 115: gb\_est46.\*
- 116: gb\_est47.\*



117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.







melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaton Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

location/Qualifiers  
1. .1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACR08K10"  
/note="end : 1ET3"  
BASE COUNT 201 a 64 c 131 g 202 t 503 others  
ORIGIN  
Query Match 0.8%; Score 56.8; DB 219; Length 1101;  
Best Local Similarity 20.3%; Pred. No. 0.0012;  
Matches 96; Conservative 174; Mismatches 202; Indels 0; Gaps 0;  
Qy 102 acatgattaccattgtttaaactgttccatcattaccataattagtaacgcattta 161  
Db 396 ATATAWAWWWTTTTTTTANAAWAAATAATTTWAAWAAWAAATAATTTWAAWAAWAA 455  
Qy 162 gtaacgcatttgtaaaatcatctgcccctttgtgtatcatatgaatgaatattat 221  
Db 456 AWAATAWTTTAAWAAAAAATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 515  
Qy 222 gattgatctgatttattgatcagaatggtgatgtatattgatgtatgctacgattgat 281  
Db 516 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 575  
Qy 282 ttgggttaactactctatgatgatattttgaaactaatctattgacttaacacc 341  
Db 576 TTYVHYVY 635  
Qy 342 atatggttataattagataatgtagctttttgtgtaaaatcaccatgcgaatttgtt 401  
Db 636 TAAAYYYYTCMYHYHYHHHAAHAAWTTTTHWTHAYHWATYHYHYHYHYHYHYHYHYHY 695  
Qy 402 ctactgttactaccatgctgaatgacgatcccaatccacagatcattcaatgaatggtg 461  
Db 696 HTCHHCYVY 755  
Qy 462 ttgtatagcaccatttaccctaatatttcaatcaaatgctatgctagcagatgatca 521  
Db 756 TWYAWAHAMWMMHHAAWAAWAAWAAWTHHYVHTTHYVHTTHYVHTTHYVHTTHYVHTTHY 815  
Qy 522 ttttttaaggtaaacaccatgaatcacatcatataaagaatcattttaaaca 573  
Db 816 WHYYTAYTCWTTHWMMWTTWHYHHHTTHWMMWTTTHWMMWTTTHWMMWTTTHWMMWTT 867

## RESULT 6

CNS02EOD/c  
LOCUS  
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 262H14 of library G from Tetraodon nigroviridis, genomic survey sequence.  
ACCESSION AL193990.1 GI:7832096  
VERSION  
KEYWORDS GSS: genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.  
REFERENCE 1 (bases 1 to 681)  
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.  
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 681)  
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.  
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence  
JOURNAL Unpublished  
AUTHORS 3 (bases 1 to 681)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
JOURNAL This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

## FEATURES

source  
1. .681  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="262H14"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG262DD07LPI-end : T7"  
BASE COUNT 74 a 105 c 170 g 282 t 50 others  
ORIGIN

Query Match 0.8%; Score 56.4; DB 220; Length 681;  
Best Local Similarity 41.1%; Pred. No. 0.0013;  
Matches 172; Conservative 26; Mismatches 221; Indels 0; Gaps 0;  
Qy 3131 accatgatcgcgttaacaaaccagtaagtggtatgatgtgaatgtggtgatgataca 3190  
Db 450 ACTTWTTCASAGACACGGAGACRATTTTWTWTATWATTTTCTCDACRACAACWACAAC 391  
Qy 3191 accattcatctaaaggcactgatgacaataaaaaacttggcgtcaaaaccaccaactg 3250  
Db 390 AACAAAAACRACWACTACAAACAAWAAACAAACGGCGGCGACAAACACGACGACGAC 331  
Qy 3251 aacaaaaaagtctaatgttaatacagcaacttaactttaatttaactctttagtgtaa 3310  
Db 330 GACACAAAAACGACRACGCGGACGACGACGACGACGACGACGACGACGACGACGAC 271  
Qy 3311 gatgccttgttaacgcgaagacatgcgcgaataatctaaacaccctagccaaaggaatt 3370  
Db 270 GACGACAAACAAACRCAACRACGACGCGGACRACAAACAAACAAACAAACAAACAA 211  
Qy 3371 cacacacaaaggcacagcacacgcctcactaaacaccttaccgtttaaaaggtagat 3430  
Db 210 AACACRCAACAAACNACGCGNACAAACAAACAAACAAACAAACAAACAAACRACRG 151  
Qy 3431 gaaaataataatgtgtgacgcgaacgcctacacgttggttcaaaagacgcaataat 3490  
Db 150 GACACAAACRACAAACAAACAAACAAACAAACAAACAAACAAACAAACRACAC 91  
Qy 3491 caagtcaacaccctaacactcaaaagtgtaaacggtcttaattattaaaccgacaaaaa 3549  
Db 90 RGCRCACAAACAAACAAACAAACRCAACRACGCGGACAAACAAACAAACAAACRACAC 32

## RESULT 7

CNS00EVL/c  
LOCUS  
DEFINITION CNS00EVL 1101 bp DNA  
GSS  
Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit



Qy	4751	accgataacaagtaactatcaagccaaaaaatgatggcscagtgtataaaacccaagaagttt	4810
Db	441	GAAATTAGGAACGCGATGATGAAGAAGACGCATGATGAAGAAGACGCAT	382
Qy	4811	gccaaagcacaaaatttgctgcgccaacgcccaaacccccagatggcacattggctcaaataaat	4870
Db	381	GATCAAGAAGACGCATGATGAAGAAGACGCATCATCAAGAAGACGCATGATGAAGAAGACGCAT	322
Qy	4871	gtcaaaatcaagtcatctaacaagaagaacaagttaaattgatgccaatataaaaggcaaggcatcaat	4930
Db	321	GATGAAGAAGACGCATGATGAAGAAGACGCATGATGAAGAAGACGCATGATGAAGAAGACGCAT	262
Qy	4931	gaagcacacgccttgtttaaaggaccttgaanaagccccttcctctatatcaacaaacccaaaac	4990
Db	261	GATGAAGAAGACGCATGATGAAGAAGACGCATGATGAAGAAGACGCATGATGAAGAAGACGCAT	202
Qy	4991	gccgcagtaactgtcggtgattttaaatgccgttgcccaa	5029
Db	201	CATCAAGAAGATGAAGATGATGAAGAATCATGATGATGATAA	163

RESULT 9

CNS000DKY/c

LOCUS

DEFINITION

CNS000DKY 928 bp DNA GSS 04-JUN-1999

Drosophila melanogaster genome survey sequence T7 end of BAC #

BAC27A24 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCESSION

AL071865

VERSION

AL071865.1 GI:4948170

KEYWORDS

GSS.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 928)

Genoscope.

REFERENCE

1 Direct Submission

AUTHORS

Submitted (02-JUN-1999), Genoscope - Centre National de Sequencage :

TITLE

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

JOURNAL

- Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila* melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org/TheBDGP/Drosophila>

melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in The Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp. the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/Drosophila\\_Bac.htm](http://bacpac.med.buffalo.edu/Drosophila_Bac.htm).

```

FEATURES
source
  1. .928
  /organism="Drosophila melanogaster"
  /db_xref="taxon:7227"
  /clone_lib="RPcI-98"
  /clone="BACR27A24"
  /note="end : T7"
BASE COUNT      262 a      70 c      84 g      321 t      191 others
ORIGIN

```

```
Query Match      0.88; Score 54; DB 219; Length 928;
Best Local Similarity 31.2%; Pred. No. 0.0062;
Matches 96; Conservative 72; Mismatches 140; Indels 0; Gaps 0;
```

[illegible]

RESULT	10
CNS00FUH/c	
LOCUS	CNS00FUH 996 bp DNA GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR31021 of RPECI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL071063
VERSION	AL071063.1 GI:4951105
KEYWORDS	GSS.
SOURCE	fruit fly, Drosophila melanogaster
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 996) Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de sequence :
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : <a href="mailto:segref@genoscope.cns.fr">segref@genoscope.cns.fr</a> - web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> )

COMMENT

web : [www.genoscope.cns.fr/](http://www.genoscope.cns.fr/)  
Determination of this BAC end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/The.BDGP/Drosophila melanogaster BAC library](http://www.fruitfly.org/The.BDGP/Drosophila%20melanogaster%20BAC%20library) was prepared by Kazutoyo Osada and Aaron Mammoser in Pietre de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's Pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES	source
found at <a href="http://bacpac.med.bu.rii.ro.edu/drosophila_bac.htm">http://bacpac.med.bu.rii.ro.edu/drosophila_bac.htm</a> .	
Location/Qualifiers	
1..996	
/organism="Drosophila melanogaster"	
/db_xref="taxon:7227"	
/clone_lib="RPCI-98"	
/clone="BACR31021"	
/note="end : TET3"	
BASE COUNT	383 a 164 c 81 g 171 t 197 others
ORIGIN	

Query Match	0.8%	Score 53.4;	DB 219;	Length 996;
Best Local Similarity	36.0%;	Pred. No. 0.0093;		
Matches 125;	Conservative	52;	Mismatches 170;	Indels 0;
				Gaps 0;



Matches	167;	Conservative	53;	Mismatches	241;	Indels	0;	Gaps	0;
---------	------	--------------	-----	------------	------	--------	----	------	----

QY	83	tgatgatttaacatgtgacatgattgaacatctgttaactactgtggccatcattacca	142
Db	1100	TTRATTTTWWTTTAAATATWAAAATAATTAATWTATTATTTAATTTWAATTTTTTTTWTATW	1041
QY	143	taatttagtaacgcattagtagcacgcatgttgaataaacatcgccgccctttatgtgat	202
Db	1040	TTWATWATTAATATWATTTTTTAAAAAATAWAAAWTTWTTTTAATTTWAATTTTTTAAATW	981
QY	203	catatagaataatatatgattgtgatcgtgatttatgtatcagaagggtgagtctatgat	262
Db	980	TTTTTWWAATTTTTTWWAATTTTWTTTAAATTTAAATTTATTTTAAATTTWAATTTWAATAA	921
QY	263	atgatgccacgsgtgttggtgggtcaacacctctgatgttgatataatttgaactaa	322
Db	920	AATATATAAAWAAATTAATTAATATTTTAAATTTAAATTTTWTATTATTTWTATTATAT	861
QY	323	tctattgacttaaatcaccatagtcgtttaatttagcataatggtaggtctttttgtaaaa	382
Db	860	TTWATTTAAATTTATAWAWTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAAAT	801
QY	383	atcacatcgcaatattgttctactgttacatcaccatgctgaatgaacgacatcccaccca	442
Db	800	WATTTTWTTTTTTTTWWATTTAATTTATATTAATTAATTAATTAATTAATTAATTAATTAAT	741
QY	443	gattcattcaagtgatgtgttgatacgcacccattaccctaatatttcaatcaaatg	502
Db	740	ATAAAAAAWAAATTTTTTTTTTTTATTAASVMTWKVWASSSCAATVTWAAAAATAATST	681
QY	503	cctatgtcagcatgcatcatcttttttaaaggtaacaccacat	543
Db	680	AMAATSVCASBSVMTTATATTTTATTAASVMTWKVWASSSCAATVTWAAAAATAATST	640

```

RESULT 12
AQ946120/c
LOCUS
DEFINITION
Sheared DNA-46J23.TR Sheared DNA Trypanosoma brucei genomic clone
Sheared DNA-46J23, DNA sequence.
ACCESSION
AQ946120
VERSION
AQ946120.1 GI:6769385
KEYWORDS
GSS.
SOURCE
Trypanosoma brucei.
ORGANISM
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE
1 (bases 1 to 641)
AUTHORS
El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
Fraser,C. and Adams,M.
TITLE
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
JOURNAL
Unpublished (1999)
COMMENT
Other_GSSs: Sheared DNA-46J23.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
http://www.tigr.org/tldb/mdb/tbdb/.
Seq primer: M13-Reverse
Class: shotgun.
```

FEATURES	Location/Qualifiers
source	1..641
	/organism="Trypanosoma brucei"
	/strain="TREU927/4 GUTat 10.1"
	/db_xref="taxon:5691"

Query Match 0.8%; Score 53; DB 219; Length 1101;  
Best Local Similarity 36.2%; Pred. No. 0.012;



```

VERSION      BF278771.1  GI:11209757
KEYWORDS
SOURCE      Gossypium arboreum.
ORGANISM    Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eustosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 780)
AUTHORS     Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
TITLE       An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
JOURNAL
COMMENT     Unpublished (2000)
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCATTATAGGG
            High quality sequence stop: 200.

FEATURES
    source   Location/Qualifiers
             1..780
             /organism="Gossypium arboreum"
             /strain="AKA"
             /cultivar="8400"
             /db_xref="taxon:29729"
             /clone="GA_Eb0033M02f"
             /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
             /tissue_type="Fibers isolated from bolls harvested 7-10
             dpa"
             /lab_host="E. coli"
             /note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT  137 a   63 c   137 g   443 t

Query Match      0.7%; Score 52; DB 146; Length 780;
Best Local Similarity 44.0%; Pred. No. 0.02;
Matches 220; Conservative 0; Mismatches 280; Indels 0; Gaps 0

QY  2990 aaagacaaaatccaacgctgcagcattatgatattattaaatcacaggctttaacctaaaa 3049
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  696 ACAACCAACACAACTAGATAATAAAAAATAAAAAATAAATAACAAAAATAAAAA 637
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  3050 aataataaacacccccattgactttgtctccacttatgacattgttgactttgccaatgac 3109
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  636 ATCAAAATAATCGAAAAAAAATATTATTAACAGGAAACAAATGAATAACATTACACTAAA 577
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  3110 aatgccaccgcgcacagtagcaccatgataccgcttaacaaacagtaagtggtatat 3169
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  576 AAAACACCCGAGACCACTATACAAAAACACCCCAATAACAAATTAATAATTAATAATAA 517
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  3170 gatgtgaatgtggatgatacaaccattctatccaacgagcactgatgacataaaaaactt 3229
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  516 AAATTAACACTGAATTAACACACACACAATAATACAACTGTACATAACAAACAATAACAAA 457
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  3230 ggcgtcaaaaccccaactgaacaaaacgaagtgttaattggttaatacagcaactaacttt 3289
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  456 AAAGCAAAACAAAAGAAAACAATAATAACACAAATTCAAATTCGCTATTAATAAACACAAA 397
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  3290 aatgttaactctagtgtgaagatgcccttgttaacgcaaaagacatcgcgaaaaatcta 3349
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  396 AAAAGAAAACAGAAATAAACAAACCTTTCACCTTTCTTCATCCCAATACAAATAAAAAAAA 337
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  3350 aacaccttagcgaaggaattcacaccaccaaggcacagcagacaccgccttacaacc 3409
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  336 AAAACCAAAATACAAAAAAGAAAAAATAAAAAAATAAAAAATAAAAAATAAAAAATAAAT 277
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  3410 ttaccggttaaaagggttagatgaaataataatgtctgatgcgcgaacgcctcacccgtg 3469
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  276 CACACCTTAAAAAATTTAAAAATATTAATAAAAAAAGAGGATAACATTAAACACATCCAAAA 217
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Qy 3470 ggtcaaaagaacgcaataa 3489  
||| | || |||  
Db 216 ACAAAAATCAGTCACATTA 197

Search completed: September 13, 2001, 03:33:34  
Job time: 53636 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 12:39:38 ; Search time 454.28 Seconds  
(without alignments)  
2905.427 Million cell updates/sec

Title: US-09-361-619-5  
Perfect score: 6972  
Sequence: 1 ccattgatggcgaggtgtg.....ccgttgacctgaggtcgac 6972

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6961	99.8	6973	1	US-08-478-370-1
2	4396.8	63.1	9542	4	US-08-968-685A-9
3	83.6	1.2	2037	4	US-08-913-942-14
4	82.8	1.2	3294	1	US-08-409-995-1
5	82.8	1.2	3294	3	US-08-685-467-1
6	82.8	1.2	3294	4	US-08-913-942-1
7	78.2	1.1	7291	4	US-08-913-942-3
8	64.8	0.9	5738	1	US-08-409-995-3
9	64.8	0.9	5738	3	US-08-685-467-3
10	61	0.9	2308	4	US-08-377-155-1
11	60.2	0.9	1797	4	US-08-377-155-12
12	60	0.9	60	1	US-08-478-370-4
13	58.6	0.8	1797	4	US-08-377-155-4
14	57	0.8	1770	4	US-08-377-155-18
15	57	0.8	1776	4	US-08-377-155-10
16	57	0.8	1776	4	US-08-377-155-20
17	57	0.8	1779	4	US-08-377-155-3
18	55.4	0.8	1785	4	US-08-377-155-6
19	55.4	0.8	1785	4	US-08-377-155-8
20	55.4	0.8	1800	4	US-08-377-155-14
21	53.8	0.8	1779	4	US-08-377-155-16
22	53.8	0.8	7218	1	US-08-232-463-14
23	50.4	0.7	2277	1	US-08-676-967-2
24	50.4	0.7	2277	1	US-08-676-974-2
25	50.4	0.7	2277	2	US-08-098-487-2
26	42.6	0.6	2947	1	US-08-457-176-1
27	42.6	0.6	2947	1	US-08-457-175-1

```

28      42      0.6      1879      3      US-08-961-083-105      Sequence 105, App
29      41.8      0.6      15222      2      US-08-801-898A-23      Sequence 23, Appl
30      41.8      0.6      15222      4      US-08-962-690-12      Sequence 12, Appl
31      41.8      0.6      15223      2      US-08-892-403A-1      Sequence 1, Appl
32      40.4      0.6      2415      1      US-08-785-428-1      Sequence 1, Appl
33      40.4      0.6      2415      2      US-08-996-797-1      Sequence 1, Appl
34      40.4      0.6      3489      2      US-08-728-323A-1      Sequence 1, Appl
35      40.4      0.6      32207      2      US-08-770-379-20      Sequence 20, Appl
36      40.4      0.6      32207      4      US-08-757-669A-13      Sequence 20, Appl
37      40      0.6      19124      2      US-08-487-826B-13      Sequence 13, Appl
38      39.8      0.6      4071      4      US-09-091-117-3      Sequence 3, Appl
39      39.8      0.6      5511      3      US-08-928-361B-2      Sequence 2, Appl
40      39.8      0.6      7334      3      US-08-928-361B-1      Sequence 1, Appl
41      39.6      0.6      1341      2      US-08-945-848-7      Sequence 7, Appl
42      39.6      0.6      2363      2      US-08-945-848-6      Sequence 6, Appl
43      39.2      0.6      819      1      US-08-320-161-3      Sequence 3, Appl
44      39.2      0.6      819      3      US-08-642-807A-18      Sequence 18, Appl
45      39.2      0.6      819      4      US-08-455-829-3      Sequence 3, Appl

```

## ALIGNMENTS

```

RESULT 1
US-08-478-370-1
; Sequence 1, Application:US/08478370
; Patent No. 5808024
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
; MEMBRANE PROTEIN OF MORAXELLA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & Mcburney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/478.370
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-502
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-478-370-1

```

Query Match 99.8%; Score 6961; DB 1; Length 6973;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6972; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

QY 1 ccattgatggcgaggtgtgctgcctgcctgatgacacccatttgc 60
   ||||||||||||||||||||||||||||||||||||||||||||

```

Db 1 CCATGGATATGGCAGGTGTGCTCGCTCGCGTATGATGGCGATCACACCCCATTTGCC 60  
Qy 61 catatctgtacgatttgacatgtgatgatatttaaacatgtgacatgatttaacattgttt 120  
Db 61 CATATCTGTACGATTTGACATGTGATGATTTTANCATGTGACATGATTTAACATTTGTTT 120  
Qy 121 aatactgttccatcatcattaccataaatttagtaacgcatattagaaacgcatattgttaaaat 180  
Db 121 AATACTGTTGCCATCATACCATTAATTTAGTACGCAATTTAGTAAAGCATTTGTAAANAT 180  
Qy 181 cattgcccctttatgtgtatcatatgaatagaatattatgatgtatctatgtattgtc 240  
Db 181 CAITTGCCCCCTTTATGCTGTATCATATGAATAGAATATTATGATTTGTATCTGATTTATGT 240  
Qy 241 atcagaatggtgatgtatgatgatgctacgaagtgtattgttgggttaatacactcatg 300  
Db 241 ATCAGAAATGTTGATGTATATGATGATGCTACGAGTTGATTTGGGTTTAACTACTATG 300  
Qy 301 atttgatatatttgaactaactctattgtacttaaatcaaccatattggtataatttagca 360  
Db 301 ATTGATATATTTTGAACCTAATCTATTGACTTAAATCACCATATGTTATATTTTAGCA 360  
Qy 361 taatggttagcttttgtaaaaatcaacatcgcaatatattgtctactgttactaccatgct 420  
Db 361 TAATGGTAGGCTTTTGTAAAAATCACATCGCAATATTGTTCTACTGTTTACTACCATGCT 420  
Qy 421 tgaatgacatcccaatcaccagattcattcaagtgtattgttggatagcaccattta 480  
Db 421 TGAATGACGATCCCATCACAGATTCATTTCANAGTGTGTTGTGTATAGCGACCATTTA 480  
Qy 481 ccctaattatttcaatcaaatgctatgtcagcatgtatcatcttttttaagggttaaacccac 540  
Db 481 CCCTAATTATTTCATCAATGCCATGTCTCAGCATGTATCATTTTAAAGGTAAACCCAC 540  
Qy 541 catgaatcaacatctataaagtcatctttaacaaagccacagggacacatttatgcagtggc 600  
Db 541 CATGAATCACATCTATAAAGTCATCTTTAAACAAAGCCACAGGCACATTTATGCGAGTGGC 600  
Qy 601 agagtacgcaaatccacacagcac- ggggggggtagctgtgtacaggggcaagtggca 659  
Db 601 AGAGTACGCCAAATCCACACAGCACGGGGGGGTAGCTGTGCTACAGGGCAAGTTGGCA 660  
Qy 660 gtgatgcactctgtagcttttgcggtattgcggtcgctgtctgtctgtatgctggtgcaaa 719  
Db 661 GTGATGACACTCTGAGCTTTTCCCGGTATTGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
Qy 720 cgtcagtggtgagtgcttatgtctcaaaaaaagataccaaacatatcgcaattgtgtgaac 779  
Db 721 CGCTCAGTGGCAGTCTTATGCTCAAAAAAAGATACCAACATATCGCAATTTGGTGAAC 780  
Qy 780 aaacccagcccaagacgtctcaggcactgccaggcgagcggtgacgagcattgctcttg 839  
Db 781 AAAACAGCCAAAGCGCTCAGGCATGCGCAAGCGGAGCGGTGATCGAGCCATTTGCTATTG 840  
Qy 840 gtgaaaaatgctaagcacaggggtcaagccatcgccatgtagtagtaataaaactg 899  
Db 841 GTGAAAATGCTAAACGACAGGGCGGTCAAGCCATCGCCATCGGTAGTAGTAAATAACTG 900  
Qy 900 Lcaatggaagcagtttggataagataggtaccgatgctacgggtcaagaggtccatcgcca 959  
Db 901 TCAATGGAAGCAGTTTGGATTAAGATAGTACCGATGCTACGGGTCAAGAGTCCATCGCCA 960  
Qy 960 tcgggtggtgatgttaaaggctagtgtgtagtgcctcgattgccaatcggttagtgatgacttac 1019  
Db 961 TCGGTGGTGATGTAAAGGCTAGTGTGTATGCTCGATTGGCATCGGTAGTATGATGACTTAC 1020  
Qy 1020 atttctgtatcagcatggttaacttaaacatccgaagggttactctgatttaacgatctta 1079  
Db 1021 ATTTCCTTGATCAGCTGTTGTTATCTTAACATCCGAAGGTACTCTGATTAAACGATCTTA 1080  
Qy 1080 ttaacggccatgcagttatataaagaaatcacgaagctcaaaagataatgatgtaaaatata 1139  
Db 1081 TTAACGGCCATCGATATTAAAGAAATACGAAGCTCAAGGATAATGATGTAAATATA 1140

Qy 1140 gacgcacaacgcgaagcggacacgcagtcactgtcagttgggagccatgtcatatgcacagg 1199  
Db 1141 GACGCACAACCAAGCGGACACGCGCAGTACTGCTAGTGGAGCCATGTCTATATGCACAGG 1200  
Qy 1200 gtcaatttttccaacgcttttgggtacacggggcaacagctaaagaatgctcctattccttggcag 1259  
Db 1201 GTCAATTTTCCAAACGCTTTGGTACACGGGCAACAGCTTAAAGTGGCTTATTCCTTTGGCAG 1260  
Qy 1260 tgggtctgcccacacagccgagggccaatctacaatcgctattggttctctgatgcaaat 1319  
Db 1261 TGGGTCCTTCCGCCACACGCGGAGGCGCAATCTTACAAATCGCTATTGTTCTGATGCAACAT 1320  
Qy 1320 ctagtctgttggagcgatagcccttgggtgcagggtactgtctcagctacagggcgagta 1379  
Db 1321 CTAGCTCGTTGGAGCGGATAGCCCTTGGTGCAGGTACTGCTGCTCAGGTACAGGGCAGTA 1380  
Qy 1380 ttgcoctagggtcaaaagtctgtgtcactcaagtgataataattcttagaccgacctata 1439  
Db 1381 TTGCCCTTAGGTCAAGGTTCCTTGTGTCACCTCAGAGTGAATAATTAATTCCTAGCCGGCTATA 1440  
Qy 1440 ccccaaatccacagcactagaccaccaagtttcaagccaccaataatacgaaggcggttc 1499  
Db 1441 CACCAAAATACCAGGCACCTAGACCCCAAGTTTCAAGCCACCAATAATAACGAAGCGGTC 1500  
Qy 1500 cactttccatgtgttagtaactctatacaacgctaaataatcaatgtcgggtcagggtgta 1559  
Db 1501 CACTTTCCATTGGTAGTAATCTCTATCAAAACGTTAAATATCAATGTGCTGCTGAGGTGTTA 1560  
Qy 1560 ataaaacgcatcggtccaatgtggcacagctagaagcggtggtgaagtgggtcgaaggagc 1619  
Db 1561 ATAAAACCGATCGGTCATGTGGCACACGCTAGAAGCGGTGAGTGAAGTGGCTTAGGAGC 1620  
Qy 1620 gtagaattacttttcagggtgatgataacagttactgacgtaaaalaggttttggataata 1679  
Db 1621 GTAGAATTACTTTTACGGGTGATGATAACAGTACTGACCTTAAATAATAGGTTTGGATAATA 1680  
Qy 1680 ctttaactattaaagggtgtgcagagacaaagcattaaaccgataataatcgggtgtg 1739  
Db 1681 CTTTAACTATTAAAGGTGGTGCAGACCAACGATTAACCGATAATAATATTCGGTGTGG 1740  
Qy 1740 taaaagggctgataatagttggtctgaaagttaaactgtctaaacactttaaacaaacttta 1799  
Db 1741 TAAAGAGCGCTGATATAGTGGTCTCAAAAGTTAAACTTCTGCTAAACTTTAAACAACTTTA 1800  
Qy 1800 ctgaggtgaaatacaactacataataatgacacacacacagttaaaggtagtagtagta 1859  
Db 1801 CTGAGGTGAATACAACCTTACATTTAAATGCCACACACAGTTAAAGGTAGGTAGTAGTA 1860  
Qy 1860 gtactacagctgaattattgagtgatgtttaacctttaccagcccaatcacaggcagtc 1919  
Db 1861 GTACTACAGCTGAATTTATTTGAGTGTAGTATTAACTTTACCCAGCCCAATACAGCAGTC 1920  
Qy 1920 aaagcacaagcaaaacgctctatggtggttaagtgtgaaagtttactataataatcgagaaa 1979  
Db 1921 AAAGCACAAGCAAAACCGTCTATGGGCTTAAATGGGGTGAAGTTTACTTAATAATGCGAAA 1980  
Qy 1980 caacagcagcaaatcggcactactcgattaccagagataaaattgcttctgctcgagatg 2039  
Db 1981 CAACAGCAGCAATCGGCACCTACTCGTATTACAGAGATAAAATTTGGCTTTGCTCGAGATC 2040  
Qy 2040 gtgatgttgatgaaaaaacaagcaccatatttggataaaaaaacacttaaaagtgggtagtg 2099  
Db 2041 GTGATGTGTGATGAAAAAAGCACCATAATTTGGATAAAAAAACAACTTAAAGTGGTAGTG 2100  
Qy 2100 ttgcaattacatagacaaatggttgaattgagtcagggttaaaaaagatcagtaactctgcca 2159  
Db 2101 TTGCAATTACCATAGACAAATGGCATTTGATCAGGTAATAAAAAAGATCAGTAACTTTGCCA 2160  
Qy 2160 aaggtacagtgcttaacgagtcggttaccatcgaaagctcaaaagccgcaagcctactt 2219  
Db 2161 AAGGTACAGTGTAAACGATGCGGTTTACCATCGACAGCTCAAAAGCCGCCCAAGCCTACTT 2220

QY 2220 taaacgagggcgctggaatcagtgatcacacacactactgaaatcatcagtgatgctaaagatg 2279  
DB 2221 TAAACGCGAGGCGCTGGCATGAGTCTACACACTACTGAAATATCATGTTGATGCTTAAGAGTG 2280  
QY 2280 gcaatgttacggccccaaacttacaacattggcgtgaaacaccacgagccttaacagtgatg 2339  
DB 2281 GCAATGTTACCGCCCACTTACACACTTGGCGTGAAACCCACCGAGCTTAACAGTGATG 2340  
QY 2340 gcaatgtgataaatttagttagtgtagtgtagtgtagtgtagtgtagtgtagtgtagtgtagtg 2399  
DB 2341 GCATAGTGTATAAATTAGTGTAAAGGTAGTGTACGACCAATAGCTTAGTTACCGCGG 2400  
QY 2400 aacatttggaagctatataaataagtcgaatgaacgagcgtgacagtgctctacaagaact 2459  
DB 2401 AACATTTGCGCAAGCTATCTAATGAAGTCAATCGAACCGGCTGACAGTGTCTACAAAGCT 2460  
QY 2460 ttaccgtttaaagaagaagacgatgatgacgccaacgcttatcacggtgcttaaaagatacga 2519  
DB 2461 TTACCGTTAAAGAAGACGATGATGACGCCAACGCTATCACCGTGGCTTAAAGATACGA 2520  
QY 2520 caaaaaatgcggcgagctcagcatctttaaactcaaaaggttaaaacggttctaaagcttg 2579  
DB 2521 CAAAAATGCGCGCAGTCAAGCATCTTAAACTCAAAAGGTAAACAGGTCTAACGGTTG 2580  
QY 2580 ctacaaaaaagatggtcaggttaccttggcttagccaagatagcgggtctgaccattg 2639  
DB 2581 CTACCAAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2640  
QY 2640 gcaaaagcaccttaaacacacagctggtgactgtttaagatacacaacgcaaaatccaag 2699  
DB 2641 GCAAAAGCACCTTAAACACACAGTGGCTTAAAGATACCAACGAAACAAATCCCAAG 2700  
QY 2700 tccgtgctaaatggcattaaatttactaatgtgaaaggtagtaacacaggttactggcattg 2759  
DB 2701 TCGTGTCTAATGGCATTAAATTACTAATGTGAATGGTAGTAATCCAGTACTGCGATTG 2760  
QY 2760 caaatccgcgtcgatccacgagataaaattggcttgcgtgctgctgctgctgctgctgctgctg 2819  
DB 2761 CAAATACCGCTCGCATTAACGAGATATAAATTTGGCTTTGGCTTTGGCTTTGGCTTTGGCTTTG 2820  
QY 2820 atacaaaaaaccttctctgatcaagaacagctacaggttggaagttaagattacca 2879  
DB 2821 ATACAAAAAACCTTATCTTGATCAAGACAAAGCTACAAAGTTGGCAATGTTAAGATTACCA 2880  
QY 2880 acactggcattaaacgaggtgtaaaagccatcacagggctgtccccaacactgctcagta 2939  
DB 2881 ACATGGCATTAACGAGGTGGTAAAGCCATCACAGGCTGTCCCCACACTGCGCTAGCA 2940  
QY 2940 ttgcgataaagtagtcgcgaacatagaaactgggcaatataacccaagaacaaagcaaat 2999  
DB 2941 TTGCCGATCAAAGTAGCGGCAACATAGAACTGGGCAATACAAATCAAGACAAAGCAAAAT 3000  
QY 3000 caacgctgcagcatataatgatattataacacaggtttaaacttaaaataaataaaca 3059  
DB 3001 CCAACGCTGCCAGCATTAATGATATTTAAATPACAGGCTTTAACTTAAATAATAAACA 3060  
QY 3060 acccattgactttgtctccattatgacattgttgactttgccaatggcaatggcaatggcaatgg 3119  
DB 3061 ACCCCATTGACTTTGTCTCCACTTATGACTTTGACTTTGCTTGGCAATGGCAATGGCAATGG 3120  
QY 3120 cgcgaacgataacccatgatccggtacacaaacagtaaaagtgatgataatgataatgataatg 3179  
DB 3121 CGCGCACAGTATCCCATGATACCGGTAAACAAACAGTAAAGTGTATATGATGTGAATG 3180  
QY 3180 tggatgatataacacattcatctaacagcactgatgacaataaaacacttggcgtcaaaa 3239  
DB 3181 TGGATGATACAAACCATTCATCTTAAACAGGACCTGATGACAATAAATAAATAAATAAATAA 3240  
QY 3240 ccaccaactgaacaaaacagtgctaatggtaatacacgcaactaactttaagttaact 3299  
DB 3241 CCACCAAACTGAACAAACAGTGTATGTTAATACAGCACTAACTTTAATGTTAACT 3300  
QY 3300 ctagtgtgaagatgcccttgttaacgccaagaacatcgccgaataatctaaacacccctag 3359

DB 3301 CTAGTGATGAAGATCCCTTGTTAACGCCAAAGACATCGCGAAATCTTAAACACCCCTAG 3360  
QY 3360 ccaaggaattcacaccaccaaaagcagcagacacccgctcaaaacctttacgctta 3419  
DB 3361 CCAAGGAATTCACACACCACCAAGGCACAGACAGACCCGCCCTACAACTTTACCGTTA 3420  
QY 3420 aaaaggtagatgaataataatgctgatgacgccaagcgcacccgctgggtgcaaaa 3479  
DB 3421 AAAAGGTAGTGAATAATAATGCTGATGACGCCAACGCCATCACCGTGGGTCAAAAGA 3480  
QY 3480 acgcaataatacaadgaacaccccttaacactcaaaaggtgaaacggtcttaataataaaa 3539  
DB 3481 AGCGAAATTAATCAAGTCAACACCTTAACACTCAAAAGGTGAAACCGTCTTAATAATAAAA 3540  
QY 3540 ccgcaaaaaatggtcaggttacctttggcattaacacacacagcgtcttaagcgcgca 3599  
DB 3541 CCGACAAAAATGTTACGGTTACCTTTGGCATTAACACCACAGCGGCTTTAAAGCGGCA 3600  
QY 3600 aaagcacccctaaacgacggtgcttctattaaaaacccactggtagcgaacaaatcc 3659  
DB 3601 AAAGCACCTTAAACGACGCTGGCTTGTCTATTAAAAACCCCACTGCTAGCGAACAATCC 3660  
QY 3660 aagtcggtgctgctgagtgaggttgcgaaggttaataataatggttctgtagtgctg 3719  
DB 3661 AAGTCGGTGTGATGGCTGAAGTTTGCCAAAGTTAATAATAATGCTGTGTAGTGTCTG 3720  
QY 3720 gcatgtatggcacaaactcgcatccacagatgaattggcttttactgggactaaatggct 3779  
DB 3721 GCATGTATGGCAACAACCTCGCATTACAGAGATGAATTTGGCTTTTACTGGGACTAATGGCT 3780  
QY 3780 cactgtataaagcaaaacccacccctaaagcaagcagcgcattaacgcaggtggttaaaaaa 3839  
DB 3781 CACTGTATAAAGCAAAACCCCACTTAAGCAAAAGACGGCATTAACGAGGTGTFAAAAGA 3840  
QY 3840 ttacaaacattcaatcagtgagatgccccaaacagccatgctgtgacagcgca 3899  
DB 3841 TTACCAACATTCATTCAGTGAGATTGCCAAACAGCCATGCTGTGACAGCGGCA 3900  
QY 3900 agattatgattaaacacggaacttgaaacaaaataacagcagctactgccaacacagcag 3959  
DB 3901 AGATTTATGATTTAAACACCGAATTTGAACAAATAATCAGCAGTACTGCCAAACACGAC 3960  
QY 3960 aaactcattacacgaattctcagtagcagatgaacaggttaataactttacggttagta 4019  
DB 3961 AAAACTCATACACGAATTTCTCAGTAGCAGATGAACAAGGTAAATAACTTTACGGTTAGTA 4020  
QY 4020 acccttactccagttatgacacactcaaaagacactctgctgctaccccttgcaggtgaaa 4079  
DB 4021 ACCCTTACTCCAGTTATGACACCTCAAAAGACCTCTGATGTCATCATCCTTTGCAGGTGAAA 4080  
QY 4080 acggcattaccacacaggttaataaaggtggtgctgctgctgctgctgctgctgctgctgctgctg 4139  
DB 4081 ACGGCTTATACCAACCAAGGTAAATAAAGGTGCTGGTGGTGGGCTTGGCCATTGACCAACCAAG 4140  
QY 4140 gcttaacacgctgaactgacgtgggtggttaataaaggtgcaaaagcattgtcattgaca 4199  
DB 4141 GCTTAAACACGCTTAAGCTGACCGTGGGTAAATAATGGCAAAAGGCATTGTTCATTGACA 4200  
QY 4200 gccaaaaatggtcaaaaataaccatcacaggactaagaacactctagcttaattgttaccatg 4259  
DB 4201 GCCAAATGGTCAAAATACCATCACAGGACTAAGCAACACTCTAGCTAATGTTACCAATG 4260  
QY 4260 ataaaggtgctgctgacacacagcaggtgcaataataatcaaaagcagacgaacaaaccc 4319  
DB 4261 ATAAAGGTAGCTACGCCACACAGACAGGCAATAATAATCAAAAGACGAGACAAACCCC 4320  
QY 4320 gtgcgcgcagcagcttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 4379  
DB 4321 GTGCCGCCAGCATTTGTTGATGTGCTAAAGCGCAGGCTTTAACTTTGCAAGCAATGTTGAAG 4380  
QY 4380 cgggtgactttgctccacttatgacacacgctcaactttgcccgtggaatgcaacacccg 4439



Db 4381 CGGTTGACTTTGTCTCCACTTATGACACGGTCAACTTTTGGCGATGGCAATGGCCACCACCG 4440  
Qy 4440 ctaaggtgacctatgatgacacaagcaaaacagtaagaatggtcttatgctatgtaaatgtgg 4499  
Db 4441 CTAAGGTGACCTATGATGACACAAAGCAAAACCAAGTAAAGTGGTCTATGATGTCAATGTGG 4500  
Qy 4500 atgatacaaccattgaagttaaagaLaaaaaacttggcgtaaaaacccacacattgacca 4559  
Db 4501 ATGATACAACCATTTGAAGTTAAAGATATAAAACTTTGGCGTAAACACCACCATTTGACCA 4560  
Qy 4560 gtactggcacaggtgtataataatttgccttaagcaatacaagctactggcgatgcgcttg 4619  
Db 4561 GTACTGGCACAGGTGCTAATAAATTTGCCCTTAAGCAATCAAGTACTGGCGATGGCGCTTG 4620  
Qy 4620 tcaagccagtgatctggtgctcatctaaacacotttatctgacgacatccaaaactggca 4679  
Db 4621 TCAAGGCCAGTGATGCTGTGCTCACTTAACACACCTTATCTGGCGACATCCAAACTGGCCA 4680  
Qy 4680 aaggggcaagccaaagcgaacactcagcaggttatgtgatgctgtagtgcgaataaagttca 4739  
Db 4681 AAGGGCAAGCCCAAGCAACTCAGCAGGCTATGCTGGATGCTGATGCAATAAAGTCA 4740  
Qy 4740 tctatgacagtacgtatacaagtaactatacaagccaaaaatgatggcacagttgataaaa 4799  
Db 4741 TCTATGACAGTACCGATTAACAGTACTATATCAAGCCAAAATGATGACAGTTGATAAAA 4800  
Qy 4800 ccaagaagttgcgaagcaaaactggtgcgcaagcccaaacccacagatggcacattgg 4859  
Db 4801 CCAAAGAGTTGCCAAGACAACACTGGTGGCCCAAGCCCAACCCAGATGGCACATTGG 4860  
Qy 4860 ctcaaatgaatgtcaaatcagttcatttaacaaagaacaagtaaatgatgccaataaaagc 4919  
Db 4861 CTCAAATGAATGTCAATCAGTCACTTAAACAAGAAACAAGTAAATGATGCCAATAAAGC 4920  
Qy 4920 aaggcatcaatgaagacaacgcttctgttaagagacttaaaagccgctcttgataaca 4979  
Db 4921 AAGGCATCAATGAAGACAACGCGCTTGTAAAGGACTTGAAGAAGCGCTCTTGATAACA 4980  
Qy 4980 aaacaaaaacgcgcagtagtaactgtggtgatttaaatgcgttgcccaaacacccgtga 5039  
Db 4981 AAACCAAAAACGCGCAGTAACTGTGGGTGATTTAAATGCCGTGGCCCAAAACCGCTGA 5040  
Qy 5040 cctttgaggggatacaggcacaaacggtcaaaaactggcgagagactttgaccatcaag 5099  
Db 5041 CCTTTGACGGGGATACAGGACACAACGGCTAAAAACTTGGCGAGACTTTGACCATCAAG 5100  
Qy 5100 gtgggcaaacagacaccaataaagtaacogataataacatcggtgtgtagcaggtactg 5159  
Db 5101 GTGGGCAACAGACACCAANTAAAGCTAACCGATATAACATCGGTGTGGTAGCAGGTACTG 5160  
Qy 5160 atggcttcaactgtcaaaactgtcaaaagacctaaccatacttaacagcgttaatgcaggtg 5219  
Db 5161 ATGGCTTCACTGTCAAACTTTGCCAAAGACCTTAACCAATCTTAACAGCGTTPAATGCAGGTG 5220  
Qy 5220 gcacaaattatgacaagggcgtctctttgtaactcaagcgtcaagccaaagcaa 5279  
Db 5221 GCACCAAAATTTGATGACAAGGCGCTGCTTTGTAGACTCAAGCGGTCAGACCAAGCAAA 5280  
Qy 5280 acacccctgtctaatgtgcaaatgggtgacctgggtggcgaaggtcaatcaagtaatgg 5339  
Db 5281 ACACCCCTGTCTAAGTGCCAAATGGCTGGACCTGGCTGGTCAAGGTCAATCAGTAATGTGG 5340  
Qy 5340 gcaagggcaaaaaagataccgacgctgccaatgtacaacagtttaaacgaagtagcgaact 5399  
Db 5341 GCAAGGCACAAAAGATACCCACGCTGCCAATGTACAACTAGTTAAACGAAGTACGCAACT 5400  
Qy 5400 tgttgggtcttggttaatgtctggttaataacgctgacgcaatcaggttaaacattgcg 5459  
Db 5401 TGTGGGTCTTGGTAAATGCTGGTAAATGATACCGCTGACGGCAATCAGGTAACATTGCCG 5460  
Qy 5460 acatcaaaaaagacccaaattcaggttcatcatctaaccgcactgtcatcaaaagcgga 5519  
Db 5461 ACATCAAAAAGACCCAAATTCAGGTTTCATCATCTAACCCGCACTGTCTCATCAAGACGCA 5520

Qy 5520 cggttacttggcgttaaaagtataaacgataccgaaaaaacttggccactggtgtatatacaag 5579  
Db 5521 CGGTACTTGGCGGTAAAGGTAAATACGATACCGAAAAACTTTGCCACTGGTGTATACAAAG 5580  
Qy 5580 tgggctgtgataaaagacggcaacgcgttaacggcgaatttaagaacaatttgggtcaaaaacc 5639  
Db 5581 TGGGCGTGTGATAAAGACGGCAACGCTAACGGCGATTAAAGCAATGTTGGGTCAAAACCC 5640  
Qy 5640 aaaaagatggcagcaaaaaaacccctgctcgccactataaaccccgaggtcagaccaact 5699  
Db 5641 AAAAAGATGGCAGCAAAAANGCCCTGCTGCCACTTATAACGCCGAGGTCAAGCAACT 5700  
Qy 5700 atttgaccacaacccccgcagagcatttgacagaaataaatagaacaaggtatccccctct 5759  
Db 5701 ATTTGACCAACAACCCCGCAGAGCCATTGACAGAATAAATGAACAAGTATCCGCTTCT 5760  
Qy 5760 tccatgtcaacatgtggcaatcaagagcctgtgtgtacaagggcgtaacggcatgtactcaa 5819  
Db 5761 TCCATGTCAACATGGCAATCAAGAGCCTGTGTACAAAGGGCTAACGGCATTTGACTCAA 5820  
Qy 5820 gtgcctcagggcaagcaactcagtggcgatagtttccagggcgaagcagatggtgaagcgc 5879  
Db 5821 GTGCCTCAGCGCAAGCACTCAGTGGCGATAGGTTTCCAGGCCAAGGCAGATGTTGAAGCCG 5880  
Qy 5880 ccgtgtccataggcagacaacccccagggcaacccaatccatcgccatcggtgtataacg 5939  
Db 5881 CCGTTGCCATAGCGACACAACCCAAAGCAGCAACCAATCCATCGCCATCGGTGTAAGC 5940  
Qy 5940 caaagccagggcgatcaatccatcgccatcggtacagggcaagcagatggttagcaggttaacg 5999  
Db 5941 CACAAGCCAGGGCGATCANTCCATCGCCATCGGTACAGCAAGTGTAAAGCTGATAACAGTTACAGTGTGGGTA 6000  
Qy 6000 actctgtgtccatcgggcagcccaagcaactgttaagctgataacaggtacaggttggtga 6059  
Db 6001 ACTCTGTGCCATCGCGACGCCCAAGCACTGTTAAGSCTGATAACAGTTACAGTGTGGGTA 6060  
Qy 6060 ataacaacagtttacgtagtgcactcaaaacgagatctcttgggtgggcaataaacaata 6119  
Db 6061 ATAACAACAGTAACTACCGATGCGCACTCAAAACCGATGCTTTGGTGGGCAATAACATCA 6120  
Qy 6120 ccgtgaccgaagtaactcggttgcttaggttcaaaactctgccaactcagtcagtcagggcacac 6179  
Db 6121 CCGTGACCGAAGTAATCTCGGTGCTTAGTTCAAACTCTGCCATCAGTCAGGCAGCAC 6180  
Qy 6180 agcagggcacaaagccaaaatactgacggcagcaggtgtaacacacacagcagggcg 6239  
Db 6181 ACGCAGGCACACAAGCCAAAATACTTGACGGCAGCAGCAGGTACACACACACAGCAGGTG 6240  
Qy 6240 caaccggtacggttaagagcttctgggcaaacgcggttgggtcggtctccgtgggtg 6299  
Db 6241 CAACCCGTACGGTTAAAGGCTTTGCTGGACAACCGCGGTGGTGGGCTCCGTTGGGTG 6300  
Qy 6300 cctcaggtgctgaacccggtatccaaaaatgtggcagcaggtgaggtcagtgccaccagca 6359  
Db 6301 CCTCAGTGTCTGAACCGCTATCCAAAATGTGGCAGCAGGTGAGGTGAGTCCACCAAGCA 6360  
Qy 6360 ccgagtcggtcaatggttagccagttgtcaaaagccaccccaaaagattgccaacgcaacca 6419  
Db 6361 CCGATGCGGTCAATGTGTAGCCAGTTGTACAAAGCCACCCAAAGCATTTGCCAACGCAACCA 6420  
Qy 6420 atgagcttgaccatgltatcccaaaaaacgaaaaaagggcaatgagggatctcatcag 6479  
Db 6421 ATGAGCTTGACCATGTATCCCAACAAAACGAAAATTAAGGCCAATGACGGGATTTTCATCAG 6480  
Qy 6480 cgaatggcagtggtggttccatggcacaagcctacattcctggcagatccatggttaccgggg 6539  
Db 6481 CGATGGCGATGGCGTCCATGCCACAAGCCCTACATTCCTGGCAGATCCATGCTTACCGGGG 6540  
Qy 6540 gtattgccaccccaacacaggttcaaggtgctggcagtgaggaactgtcgaagctctcgata 6599  
Db 6541 GTATTGCCACCCCAACGGTCAAGGTGCGGTGGCAGTGGGACTGTGGAAGCTGTGCGGATA 6600





Db 984 CCGTTGGTGAAGTACTCAAGTTAAATTCGAATCGAGGTATTCGCCCTAGGTTTGGTTCTC 1043  
QY 1401 ttgtcactcagagtgataaat-----tctagacggcctatacacca---a 1445  
Db 1044 AGGTCCTTCAGAGGATATGATGAATGCAGCAAAATGACGGGCTATGCACGAGATG 1103  
QY 1446 ataccaggcactagaccccaagtttcaagccac---caataatcgaaggcggtccac 1502  
Db 1104 ATAACCGACCAATAGACAACCGGTATAAGCCACCCTTCAAGAATGTTGCTACGGATGAT 1163  
QY 1503 ttccattggtagta-----actctatcaaacgtaaaatcatcaatgcggtg 1550  
Db 1164 TTTCATTTGTTAGTAAATGGGAATGACAGATGACAGCGGTAAATCATCAATGTCGGTG 1223  
QY 1551 cagggtgtaataaaacgcatgoggttcaatgtagtggtgacagctagaagcgggtggaagtgg 1610  
Db 1224 CAGGTTCTGGGGTACCGATGCGGTCAATGTGGCACAGTTAAAGAGCGGTGACGGCTGG 1283  
QY 1611 ctaagagcgtagaattactttcagggtgtagatacagctactgacgttaaaaaataggtt 1670  
Db 1284 CTA---ATCGTCAATTTACTTTTAAAGGTGATGATAGCAATATAGAGTAGAATAAGGTT 1340  
QY 1671 tggataactttaactattaaagggtggtgcagagacccacgcatcattaaoccgataata 1730  
Db 1341 TGGCAAGACTTTAACTATCACAGGTGGTGCACAGACCGCATTAACCGATCATAAACA 1400  
QY 1731 tgggtggtgtaaaagggtgataatagtggtctgaagttaaaactgctaaacttaaa 1790  
Db 1401 TCGGTGTGGTACAAAATGGCGA-----TGGTCTGAAGTTCAACTTGTGTAACATTTAA 1454  
QY 1791 acaacttactgaggtgaaTacaactacattaaatgcacacacacagtttaaggttaggta 1850  
Db 1455 CCAGCCTTAAATGGTTTACCACTGAAACCTTACCGCCACGAGAAAGTTACCGTAGGCA 1514  
QY 1851 gtagtagtagtactacagct-----gaattcattgagt 1882  
Db 1515 AAACCCGCTTACCACAGATAAAATTTGGTTTACCAATGATATGAATGGCATTTGATCAAA 1574  
QY 1883 gatagtttaacctttaccagcc-----caatcagcagctcaaaagcacaagcaaaa 1934  
Db 1575 GCAACCTTATFTCTGATAAGACACTCGCAATTCATGACAGTGGTCAAAAGATTACCAAAC 1634  
QY 1935 ccgtctatggtgtaattggtggaagtttactaataatgcagaacacacagcgcaatcg 1994  
Db 1635 TTAGTGTGGTGTAGTAGACGATCGGCAACTTATGACAGCTTAAATAAAGTTAAACC 1694  
QY 1995 gcaactactgtattac-----cagagataaaattggctttgctcgagatggtg 2042  
Db 1695 AAACCGGTGAAGTGCTCTACAAACCTTTACCGTTAAATAAGGTAGATAAAATGGTAATG 1754  
QY 2043 atgttgatgaaaaaacagcaccaatatttggataaaaaaacacttaaaagtgggtagtggtg 2102  
Db 1755 ATGCTAATGACAGCAAAATCATCCGCTGGGTAAATAAACAACACGAGCTACTCAAG 1814  
QY 2103 caattaccata-----gacaatggcattgtagcaggttaataaaagaatcagta 2150  
Db 1815 TCAACACCTTAAACTCAAAGGTGAAACGGTGTGTATGTATACAACCGCAACAAATGGTA 1874  
QY 2151 atcttgcaaggtagcagtgctaacggtggttacatcgcaacagctcaaacggccca 2210  
Db 1875 CAGTTTACCTTTGGGCTTAAACCAAAATTAACGGTCTGACCGTTGGCAACAGCACCTTAACA 1934  
QY 2211 agcctactttaaacgacggcgtggcactcaggtcaccacttactgaaatcagttgagt 2270  
Db 1935 ACGATGGCTTATCTTTAAACACCAATAGTAAACAACAAATCCAAAGTCGGTGTCTGATG 1994  
QY 2271 ctaagagtggaatttaccgcccccaacttaacaactggcgt-----gaaacacccg 2324  
Db 1995 GCATTTACATTTACTGATATACGAATAGTAGCCAGGTGCTGGCATTTGAATAATACCACTC 2054  
QY 2325 agcttaacagtgatggcactagtgataaaatttagttaggttagtgacgaacata 2384

Db 2055 GCATTACCAGACAGCGTATTTGGTTTGTCTAATAATACTGGTTTCAATGGATGCAAAACAAC 2114  
QY 2385 gcttagttacc-----gcgaacatttggccaagctatctataatg 2423  
Db 2115 CCGGCTTAACCCCAACTGGCATTTAACGCGAGTGGTAAAGAGCTGACCAATGTCCAATCTG 2174  
QY 2424 aagtcactcgaacggctgacagtgctctctacaagcttt----- 2461  
Db 2175 CCATTAACCCCTGCTACCAATATGGTGGCAGCTAGACTTTATTAAGCCGCCCTAAGCAGCTGCTA 2234  
QY 2462 -----accgttaagaagaagacagatgtagcgcacacg 2495  
Db 2235 ATACCGAAAAATCAGGCTCVCYGCCGCCACCATTAAGACTTTATACACCTATACAAAGTAC 2294  
QY 2496 ctatcacgctggctaaagatacgcacaaaaatgcccggcgca-----gtcagcatcttaa 2549  
Db 2295 CGCTGACCTTTGCAGGTGATACAGGTCTCTAATGTACCACAAAAACTGGGCGAGATTTTAA 2354  
QY 2550 aactcaagggttaaaacggctctaaagggtgtgtctaccacaaagaagatggtacgggttacctt 2609  
Db 2355 AGGTTTAAAGGTGGTAAACACACAGCTGATGATTTTAAACCAAAATTAACATCGGTGTGGTGG 2414  
QY 2610 ggcttagccaagatagcggtctgaccatt----- 2638  
Db 2415 CTGATAGTAGACCGATTAATAGCTTAACCGTTAAACCTTTAAACCTTTAAACCGATCTTGATG 2474  
QY 2639 -----ggcaaaagcacccctaaacacacgacgatggcttga 2669  
Db 2475 CGGTTAATACTTAAACCCCTTACTGCGCAGCGATAAAAGTTTACCGTAGACAGTGGCAACAACA 2534  
QY 2670 ctggttaagatataccaa-----cgaaacaaatccaaatcggtgctaatg 2711  
Db 2535 CGGCTAAGCTACAAAATGGTGATTTAACTTTAGCAAAACAAAATACAGGTGCTACCCCTG 2594  
QY 2712 gcatlaaatttactaatgtgaatg-----tagtaaatccaggtta 2750  
Db 2595 CCACCAACAGCAAAACCATCTATGCGCTTGAGCTTGAAGTTTACTGATAACAATGGTA 2654  
QY 2751 ctggcattgcaaataccgctcgcattaccagagataaaatgtgcttctgctgtatg 2810  
Db 2655 TAGCATTGACGGCAGCACTTACATCACCAAGACAAAAGTTGGCTTTGCTTAAGCAAGATG 2714  
QY 2811 gtgcagtgatacaaaacaaaccttcttgatcaagacaaagctacaagttgccaatgta 2870  
Db 2715 GTTCACTGTATAAGCAAAACCTTATCTTGATAAGGCACAAAGCTTAAAGTGGGTGAAGTTG 2774  
QY 2871 agattaccacactggcatttaacgcagggtgggttaaagccatcacagggtgttcccccaaac 2930  
Db 2775 AGATTACCACCAACAGCGCATTAATGCAAGTGGTAAAGCCATCACAGGACTTAAGCAATACCC 2834  
QY 2931 tgcctageattgccgatcaagtagccgaacatagaactgggcaatacaatcccaagca 2990  
Db 2835 TAACCGATGCCACCAAGCAACACAGGCGATGTACTCAATTTGGGTATCGTTGATAGTA 2894  
QY 2991 aagacaaaatcccaacgctgcagcatttaataatgatatataatcaacagctttaaaccctaaaaa 3050  
Db 2895 CTGACAAAACCCGTGCGCGCAGCATTTGGTGATGTGCTTAAACGCGAGGCTTTAACTTAAAAA 2954  
QY 3051 ataatacaaacccccattgactttgtctccacttaagacattgttgaactttgccaatggca 3110  
Db 2955 ATATGGTGCACGCGCAAGACCTTTGTCTCCACTTATGACACTGTGTGATTTATCAATTTGGCA 3014  
QY 3111 atgcccaccacccacagtaaacccatgataccgctcaacaaacacagtaaaagtgggtatag 3170  
Db 3015 ATGCCACACCGGTAAAGTCACCTTATGA-----TGGCAAAAGCCACTAAGTGGCGGTATG 3068  
QY 3171 atgtgaatgtggatagatacaaacatttcatctaaacaggtcagtgatgacataaaaa----aac 3227  
Db 3069 ATGCTAATGTGGATGTGTACAAACCATTCATCTAACAGCGCGCTGATGCAATTAATAAACCAAA 3128  
QY 3228 ttggcgtcaaaacacacacaaactgaacaaaacagtgctaatgtgataacagcaactaac 3287  
Db 3129 TTGGCGTAAAAACACCAACACTGACCAAAACAGATGCTTAAGGTGATTAAGCAATTAAC 3188

Qy	3288	ttaa	gt	taactc	tagt	gag	atg	ccctt	gtt	aaacg	ccaa	agacat	cgcgcg	aaaaac	3347								
Db	3189	TTAG	TGTTAACT	CTCGT	GTATG	ATCA	AAAGCCCT	TATTAA	CGCCAA	AGACAT	CGCG	CAATC	3248										
Qy	3348	taaac	accct	tagc	caag	gaaa	ttc	acac	ccca	aaag	ccac	ag	cag	cag	ccct	caaaa	3407						
Db	3249	TAAAC	ACCC	TAGT	GGT	GAA	TTCC	ACAC	ACAN	AGGC	CAC	AGAC	CAC	CGCCT	TAC	AAA	3308						
Qy	3408	cc	tttacc	gtt	taaaa	agg	tag	atg	aaaa	---ta	aatg	ctg	atg	cgc	caaac	gcgc	atca	3464					
Db	3309	CCTTT	CAAGT	CAAAA	AAAGT	CAAAA	AAAA	TTGGT	GAT	GAT	GA	TAA	TAG	CGCT	GC	TAC	ACCATCA	3368					
Qy	3465	ccg	tg	gggt	caaa	agac	gc	caaa	ta	at	caag	tc	aac	ccct	taaac	act	caaa	gggtg	aaaaacg	3524			
Db	3369	CCGT	GGGTAA	AGAT	GC	AAAA	ACCA	ATCA	AGT	CAAC	ACCC	TAAAA	CTCAA	AGG	CTAAAA	ACG	3428						
Qy	3525	gtc	t	ta	at	taaaa	ccgc	aaaa	atg	gt	c	ggt	t	ac	ctt	tg	gc	at	taac	ccac	caagc	3584	
Db	3429	GTCT	TG	NTAT	T	C	N	A	C	C	A	A	T	A	A	G	T	T	A	A	C	C	3488
Qy	3585	gtc	t	taaa	ccgc	gcg	aaaa	agc	acc	---c	t	aaac	gc	ag	ctg	ggc	tt	g	ct	t	taaaa	ccccca	3641
Db	3489	GTCT	TAA	AGCGG	CAAC	AAAC	AC	CACT	CT	TAA	ACA	CAAT	GTG	CTT	GTCT	TATT	TAAAA	ACAC	CCG	3548			
Qy	3642	ct	g	t	a	g	c	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	3701
Db	3549	CTGGT	AAC	CAAC	AAAT	TCC	AAAGT	CGGT	GTG	TGA	TGG	CGT	GAA	GTT	TGCC	AAAG	CT	---TAA	TA	3605			
Qy	3702	at	g	g	t	g	t	a	g	t	g	c	t	g	c	a	a	a	a	a	a	a	3761
Db	3606	ATGGT	GT	T	GT	AGG	TCT	G	C														

QY	4362	tqcaaggcaatggtgaagcggttgactcttgtctccaacttaigacacogtcaactttgccc	4421
DB	4266	TCGCAAGGCATGGTGAAGCGTTGACTTTGTCTCCACTTATATGACTGTCAACTTTGCCG	4325
QY	4422	atggcaatgccaccaccgctaaggtgacctaigtacacaaacgaaacacgataaagtgg	4481
DB	4326	ATGGCANTGCCACACCGGTNAGGTGACCTTATGATGACACACGAAACACGTAAGTGG	4385
QY	4482	tctatgatgtcaatgtggatgatacaacccattgaagttaaagataaaaaacttggcgtaa	4541
DB	4386	TCTATGATGTCAATGTGGTATGATACACCATTTCAAGTTAAAGATAAAAAACTTTGGCGTAA	4445
QY	4542	aaacacacacattgaccagttactggtgcacaggtgctaatataatttgccctaagcaatcaag	4601
DB	4446	AAACACCAACATTGACACGACTTGGCACAGGTGCTAATAAATTTGGCCTAAGCAATCAAG	4505
QY	4602	ctactggcgatcgctcttcaaggccagtgatatctgtctcatcttaaacacaccttatctg	4661
DB	4506	CTACTGGCGATGGCTTGTCAAGCCACGTATATCGTTGCTCATCTAATACACCTTATCTG	4565
QY	4662	gggacatccaaactgccaaaggggcgaagccaaagcgaacaaactcagcaggtctatgtggatg	4721
DB	4566	GGGACATCCAACTGCCAAGGGGCAAGCCAAAGCAACAGCTCAGCAGGCTATGTGGATG	4625
QY	4722	ctbtagtgcaatgaaggtcatctctatgacagtaccgataaacagttactactcaagccaaaatg	4781
DB	4626	CTGATGGCAATTAAGGTCTACTATGACAGTACCGATACCAAGTACTATCAAGCCCAAAAATG	4685
QY	4782	atggcacagtgtataaaaccaaagaagtgtgcaaaagacaaaactggctgcgcccagccccaaa	4841
DB	4686	ATGGCACAGTTGATAAAACCAAGAAGTTGCCAAAGACAAACTGGTGCCGCCAAGCCCCAAA	4745
QY	4842	ccccagatggacattggctcaaatgaaatgtcgaattcagtcattacaagaagaacaagtaa	4901
DB	4746	CCCCAGATGSCACATTGGCTCAAAATGAATGTCAATCAGTCAATTACAAAGAACACAGTAA	4805
QY	4902	atgatgccataaaaaagcaaggcatcaatgaagacaacgcctttgttaaaggacttgaaa	4961
DB	4806	ATGATGCCAATTAAGACGAAGGCATCAATGAAGACAAACGCCTTTGTTAAAGGACTTGAAA	4865
QY	4962	aagcgcttctgatacaaaaaccaaagccagtcgaactgtgggtgatttaaatgcg	5021
DB	4866	AAGCGCTTCTGATACAAACCAAAACGCCGACGTAACTGTGGGTGATTTAAATGCCG	4925
QY	5022	ttgccaaacacgcgtgacctttgcggggatcacaggcaacgcgctaaaaaactggcg	5081
DB	4926	TTGGCCAAACACCGCTGACCTTTGCAGGGGATACAGGCACACCGCTTAAAAACTGGCG	4985
QY	5082	agacttggacctcaaaagtgggcaacagacacccaataagctaaocgataataacatcg	5141
DB	4986	AGACTTTGACCATCAAAAGTGGGCAACACAGACACCAATAAGCTTAACCGATAATAACATCG	5045
QY	5142	gtgtgttagcaggtactgatggcttcactgtcaacttgccaaagacctaaccaatotta	5201
DB	5046	GTGTGTGTAGCAGTACTGATGGCTTCACTGTCAAACTTGGCAAGACCTTAACCAATCTTA	5105
QY	5202	acagcgttaatgcaggtggcaccaaaattgtacaaaagcgctgtcttttgttagactcaa	5261
DB	5106	ACAGCGTTAATGCAGGTGGTACCAAAATGATGACAAAGCGGTGCTTTTGTAGACTCAA	5165
QY	5262	ggcgtaagccaaagcaaacaccccttgcctaagtgccaaatgggctggacctgggtggca	5321
DB	5166	GGCGTCAAGCCAAAGCAACACCCCTGTGCTAAGTGCCAATGGCTGGACTGGGTGGCA	5225
QY	5322	aggtcatcagtaatgtgggcaaaaggcacaagaataccgcgtgcgaatgtacaaacagt	5381
DB	5226	AGGTCAATGACAAATGTGGCAAGAGGCAAAAGACACCCGACGTGCCAATGTATCAACAGT	5285
QY	5382	taaacgaagtacgcaacttgttgggtcttggtaatgctgtgtaatgataaacgctgacggca	5441
DB	5286	TAAACCAAGTACGCAACTTTGTTGGGTCTTGGTTAATGCTGTTAATGATTAACGCTGACGGCA	5345
QY	5442	atcaggtaaacattgcccgcacatcaaaaaagaccaaaattcaggttcatcatcaaacgcga	5501

Db 5346 ATCAGTAAACATTTGCCGACATCAAAAAAGACCACCAATTTTCAGTTTCATCATCTAACCGCA 5405  
Qy 5502 ctgtcatcaaaagcagcagcagctacttggcggttaaaggttaataacagataccgcaaaacttg 5561  
Db 5406 CTGTCTATCAAGCAGCAGCAGCTACTTTGGCGGTAAAGGTAATAACGATACCGGAAACATTG 5465  
Qy 5562 ccaactggtgatacaagtgggcggtgataaaagacgagcaacgtaacgagcgtatttaagca 5621  
Db 5466 CCACTGGTGTACANGTGGCGCTGGNTAAAGACGGCAACGCTAACGGCGGATTAAAGCA 5525  
Qy 5622 atgtttgggtcaaaaccccaaaagatggcagcaaaagccctgctgcgcacttataacg 5681  
Db 5526 ATGTTTGGGTCAAAACCCAAAAAGATGGCAGCAAAAAAGCCCTGCTGCGCCACTTATAACG 5585  
Qy 5682 ccgcaggtcagaccaactatttgaccaacaaccccgacagaagccattgacagaataaag 5741  
Db 5586 CCAGCAGGTACAGCAACTATTTCACCAACAACCCCGCAGAACCAATTGACGAATAAATG 5645  
Qy 5742 acaaggtatccgcttcttcctcaatgtaacagatggcaatcaagagcctgtggtacagggc 5801  
Db 5646 AACAGGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCCTGTGGTACAGGGC 5705  
Qy 5802 gtaacggcatgtactcaagtgcctcgggcaagcaactcagtggtgagtaggtttccaggcca 5861  
Db 5706 GTAACGGCATTTGACTCAAGTGCCTCAGGCAAGCACTCAGTGGCGATAGGTTTCCAGGCCA 5765  
Qy 5862 agcaaatggtgaagccgctgtgcatagcagacaaacccaagcagcaacaaatcca 5921  
Db 5766 AGCAGATGTTGAAGCCGCCCTTGCCTAGCCAGCAACAAACCCCAAGCAGCAACCAATCCA 5825  
Qy 5922 tcgcatcggtgataacgcaaaagccagcggtgataatccatcgcctcagtggtacaggca 5981  
Db 5826 TCGCCATCGGTGATAAGCCACAAGCCACAGCGCATCAATCCATCGCCATCGGTACAGGCA 5885  
Qy 5982 atgtgtagcaggtgaagcaactctgtgcatcggcgcaacaaacactgttaagctgata 6041  
Db 5886 ATGTGTAGCAGGTGAAGCATCTGTGTGCCATCGCGCAACCAACCACTGTTAAAGCGTGATA 5945  
Qy 6042 acagttacagtggtggttaataacaaacagtttacccgatgcactcaaacagatgctttg 6101  
Db 5946 ACAGTTACAGTGTGGTGAATAACAAACCAAGTTTACCGATGCCACTAAACCGATGCTTTG 6005  
Qy 6102 gtgtgggcaataacatacccgtagcagaaagtaactcgtgtgcttaggttcaaacctctg 6161  
Db 6006 GTGTGGGCAATAACATCACCGTACCGAAAGTAAGTCTGCTGTAGGTTCAAACTCTG 6065  
Qy 6162 ccatacgtcagcgacacacgagcgacacaaagccaaatactgacgacacagcaggtga 6221  
Db 6066 CCATCAGTGCAGGACACACGCGAGGACACAAAGCCAAATACTGACGGCAGCAGGTA 6125  
Qy 6222 caaccacacagcaggttgcaaccggttacggttaaaggcttctgtgacaaacgcggttg 6281  
Db 6126 CAACCCACACAGCAGGTGAACCGGTACCGTAAAGGCTTTGCTGGCAAAACGCGGTTG 6185  
Qy 6282 gtgcggtctcgtgtggtgctcaggtcgtgaagcgtatccaaaatgtggcagcaggtg 6341  
Db 6186 GTGCGGTCTCGTGGGTGCTCAGGTGCTGAACGCCGTATCCAAATGTGGCAGCAGGTG 6245  
Qy 6342 aggtcagtgccaccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6401  
Db 6246 AGGTGAGTGCACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 6305  
Qy 6402 gcatgtccacgcaaccaaagatggtgaccatgataccaccacaaacgaaataaaggcca 6461  
Db 6306 GCATTGGCAACGCAACCAATGAGTTGACCATGTTACCATGTTACCAACCAACCAATTAAGCCA 6365  
Qy 6462 atgcagggattctatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 6521  
Db 6366 ATGAGGATTTATACGAGTGGCGTATGCGTCCATGCGCACAGCCCTACATTCCTGGCA 6425  
Qy 6522 gatccatggttacgggggttatgtccaccccaacacgctcaggtgcggtgagcagtgagac 6581

Db 6426 GATCATGTTTACCGGGGGTATTTGCCACCCACAAACGGTCAAGTGGCGTGGAGTGGGAC 6485  
Qy 6582 tgcgaagctgcgagataatggtcaatgggtatttataaatacaatgggttcagccgatacccc 6641  
Db 6486 TGTCAAGCTGTGCGATATGTCATGCGTAAATGGGTATTTAAATCAATGGTTTACGCCGATACCC 6545  
Qy 6642 aaggccatgtaggggcgagtggtgaggttttcaacttttaagccataaatacgcaaga 6701  
Db 6546 AAGGCCATGTAGGGCGGCGAGTTGGTGCAGGTTTTCACCTTTAAGCCATAAATCGCAAGA 6605  
Qy 6702 ttttacttaaaatacaatctccacatagttgataaaaacacacacacacacacacacacac 6761  
Db 6606 TTTTACTTAAAAATCAATCTCACCATAGTTGTATAAAAACAGCATCAGCATCAGTCATAT 6665  
Qy 6762 actgatgctgattgttttttatacacttaaacacacacacacacacacacacacacacac 6821  
Db 6666 ACTGATGCTGATGTTTATATCATTAAACCATTTTACCGCTCAAGTGAATATCTTTCAC 6725  
Qy 6822 catgaccaatacgcactgactgactaggttaaaccttattgagtaaatatttatcaatgatt 6881  
Db 6726 CATGACCAAAATGCCATTTGATCATAGTAGTAACCTTATTTGAGTAATTTTATCAATGAGTT 6785  
Qy 6882 gttagataggttataaattgtccattgacccaaataagacccgatttatcccgaaaaatt 6941  
Db 6786 GTTAGATATGGTTAAATTTGTCCTCATTTGACCAAAATTTACCGATTTATCCCGAAAAATTT 6845  
Qy 6942 ctgattatgatccgttgacctgcaggtc 6969  
Db 6846 CTGATTATGATCATTCTTTTCATAAAATTC 6873

RESULT 3  
US-08-913-942-14  
; Sequence 14, Application US/08913942  
; Patent No. 6200578  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme, Joseph  
; APPLICANT: Barenkamp, Stephen J.  
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Flehr Hohbach Test Albritton & Herbert LLP  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,942  
; FILING DATE: 29-DEC-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/409,995  
; FILING DATE: 24-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/4031  
; FILING DATE: 22-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Vance, Dolly A.  
; REGISTRATION NUMBER: 39,054  
; REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2037 base pairs





Db 3223 ATTATTCCTGTGTCAGGCACCAACCAATAGTCAAGGTAAACAGCGTGTGCAGCAGGTGTT 3282  
Qy 6671 ggtttcaact 6680  
Db 3283 GGTACCAGT 3292

RESULT 7  
US-08-913-942-3  
; Sequence 3, Application US/08913942  
; Patent No. 6200578  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme, Joseph  
; APPLICANT: Barenkamp, Stephen J.  
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,942  
; FILING DATE: 29-DEC-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION NUMBER:  
; APPLICATION NUMBER: US 08/409,995  
; FILING DATE: 24-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/4031  
; FILING DATE: 22-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Vance, Dolly A.  
; REGISTRATION NUMBER: 39,054  
; REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7291 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 163..7221  
; US-08-913-942-3

Query Match 1.18; Score 78.2; DB 4; Length 7291;  
Best Local Similarity 52.8%; Pred. No. 1.6e-11;  
Matches 198; Conservative 0; Mismatches 168; Indels 9; Gaps 1;

Qy 6320 atccaaatgtggcagcaggtgaggtcagtcagtcaccagcagcagcagtcggtcaatggtagc 6379  
Db 6850 ATTGACAATGTGGCTAATGGCGAAATTTCTGCCACTTCCACCGATGCCGATTAAACGGAAGT 6909  
Qy 6380 cagttgtac-----aaagccaccacaaagcattgccaacgcaacaaatgagcttgac 6430  
Db 6910 CAGTTGTATGCCGTGGCAAAAGGGGTAAACAACCTTGTGTGGACAAGTCAATAATCTTGAG 6969  
Qy 6431 catcglatccaccacaaacgaataaagcccaatgcagggatttcacatgcagcagcgatg 6490

Db 6970 GGCAAAAGTGAATAAGTGGGCAAAAGCTGCAGATGCAGGTACAGCAAGTGCATTAGCGGCT 7029  
Qy 6491 gcgctccatgcacaagcctcatcttcctggcagatccatggttaccgggggtatttccacc 6550  
Db 7030 TCACAGTTACCACAGCCACTATGCCAGGTAAATCAATGGTTGCTATTTCGGGGAAGTAGT 7089  
Qy 6551 cacaacgggtcaaggtgcggtggcagtggtgagctgtcgaagctgtcggaatggtcaatgg 6610  
Db 7090 TATCAAGTCAAAATGGTTAGCTATCGGGTATCAAGAAATTCGATAATGGCAAGTG 7149  
Qy 6611 gattttaaatcaatggttcagccgatacccaagcgaatgtaggggcgagttggtgca 6670  
Db 7150 ATTATTCGCTTGTGAGGCACCAACCAATAGTCAAGGTAAACAGCGTGTGCAGCAGGTGTT 7209  
Qy 6671 ggtttcactttta 6685  
Db 7210 GGTACCAGTGGTAA 7224

RESULT 8  
US-08-409-995-3  
; Sequence 3, Application US/08409995  
; Patent No. 5646259  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen I.  
; APPLICANT: St. Geme III, Joseph W.  
; TITLE OF INVENTION: Haemophilus Adhesion Proteins  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/409,995  
; FILING DATE: 24-MAR-1995  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-61053/RFT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5738 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; US-08-409-995-3

Query Match 0.9%; Score 64.8; DB 1; Length 5738;  
Best Local Similarity 66.4%; Pred. No. 8.6e-08;  
Matches 93; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 4311 acaaaacccgtgcgcccagcagctgttgatgtgtactaagcgcagcgcttacttgcaggca 4370  
Db 2621 AAAAATCCAATGCAGCAAGTAGTTGAAGATGTATTCGGCGCAGGTGGAATATTCAAGGTA 2680  
Qy 4371 atggtgaagcgtgtgactttgtctccacttatgacacgcgtcaactttgccgatggcaatg 4430  
Db 2681 ATGGTAATAATGTTGATTATGTAGCGACGTATGACACAGTAACATTTACCGATGACAGCA 2740















GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 12:39:38 ; Search time 33595.7 Seconds  
(without alignments)  
3209.970 Million cell updates/sec

Title: us-09-361-619-5  
Perfect score: 6972  
Sequence: 1 ccatggatatgggcaggtgt.....ccgttgacctgcaggtgcac 6972

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_ba3: \*  
4: gb\_in1: \*  
5: gb\_in2: \*  
6: gb\_in3: \*  
7: gb\_om: \*  
8: gb\_ov: \*  
9: gb\_pat1: \*  
10: gb\_pat2: \*  
11: gb\_ph: \*  
12: gb\_pl1: \*  
13: gb\_pl2: \*  
14: gb\_pl3: \*  
15: gb\_pl4: \*  
16: em\_ba1: \*  
17: em\_ba2: \*  
18: em\_fun: \*  
19: em\_htgo\_hum: \*  
20: em\_htgo\_inv: \*  
21: em\_htgo\_rod: \*  
22: em\_htg\_hum1: \*  
23: em\_htg\_hum2: \*  
24: em\_htg\_hum3: \*  
25: em\_htg\_hum4: \*  
26: em\_htg\_hum5: \*  
27: em\_htg\_hum6: \*  
28: em\_htg\_hum7: \*  
29: em\_htg\_hum8: \*  
30: em\_htg\_inv1: \*  
31: em\_htg\_inv2: \*  
32: em\_htg\_other: \*  
33: em\_htg\_rod: \*  
34: em\_hum1: \*  
35: em\_hum2: \*  
36: em\_hum3: \*  
37: em\_hum4: \*  
38: em\_hum5: \*  
39: em\_hum6: \*  
40: em\_hum7: \*  
41: em\_in: \*  
42: em\_om: \*  
43: em\_or: \*

44: em\_ov: \*  
45: em\_pat: \*  
46: em\_ph: \*  
47: em\_pl: \*  
48: em\_ro: \*  
49: em\_sts: \*  
50: em\_sy: \*  
51: em\_un: \*  
52: em\_vi: \*  
53: gb\_sts1: \*  
54: gb\_sts2: \*  
55: gb\_sts3: \*  
56: gb\_sy: \*  
57: gb\_un: \*  
58: gb\_vil: \*  
59: gb\_vil2: \*  
60: gb\_vil3: \*  
61: gb\_vil4: \*  
62: gb\_vil5: \*  
63: gb\_vil6: \*  
64: gb\_vil7: \*  
65: gb\_vil8: \*  
66: gb\_vil9: \*  
67: gb\_vil10: \*  
68: gb\_vil11: \*  
69: gb\_vil12: \*  
70: gb\_vil13: \*  
71: gb\_vil14: \*  
72: gb\_vil15: \*  
73: gb\_vil16: \*  
74: gb\_vil17: \*  
75: gb\_vil18: \*  
76: gb\_vil19: \*  
77: gb\_vil20: \*  
78: gb\_vil21: \*  
79: gb\_vil22: \*  
80: gb\_vil23: \*  
81: gb\_vil24: \*  
82: gb\_vil25: \*  
83: gb\_vil26: \*  
84: gb\_vil27: \*  
85: gb\_vil28: \*  
86: gb\_vil29: \*  
87: gb\_vil30: \*  
88: gb\_vil31: \*  
89: gb\_vil32: \*  
90: gb\_vil33: \*  
91: gb\_vil34: \*  
92: gb\_vil35: \*  
93: gb\_vil36: \*  
94: gb\_vil37: \*  
95: gb\_vil38: \*  
96: gb\_vil39: \*  
97: gb\_vil40: \*  
98: em\_ba3: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB ID	Description
1	6972	100.0	6972	10	AX079917 Sequence
2	6961	99.8	6973	9	AR040716 Sequence
3	6961	99.8	6973	10	AX079913 Sequence
4	6258	89.8	6259	10	AX079924 Sequence
5	6141	88.1	6141	10	AX079918 Sequence
6	5976	85.7	5976	10	AX079914 Sequence
7	4671.2	67.0	62909	10	AX067457 Sequence
8	4101.6	58.8	6159	10	AX079920 Sequence



QY 1141 acgcacaaaccgcaagggacacccagctactgcagtgaggagccatgtcatatgcacaggg 1200  
 Db 1141 ACGCACAACCGCAAGCGGACACGCCAGTACTGCGAGTGGGAGCCATGTCTATATGCACAGGG 1200  
 QY 1201 tcaattttccaaagccttttggtacacgggcaacagctaaagtgccatttccttggcagt 1260  
 Db 1201 TCATTTTTCCAAAGCCCTTTGGTACACGGGCAACAGCTTAAAGTGCCTTATTTCTTTGGCAGT 1260  
 QY 1261 gggcttggccacacgaggggccaactcaaatcgctatggttctgtatgcacacatc 1320  
 Db 1261 GGCTCTGGCCGCACAGCGGAGGCCAATCTACAATCGCTATTGGTCTCATGCACATC 1320  
 QY 1321 tagctcgttggggcgatagcccttgggtgcaggtactcgtgcagctacagggcgagt 1380  
 Db 1321 TAGCTCGTTGGGAGCATAGCCCTTGGTGCAGGTACTCGTGCCTCAGCTACAGGCGCAGTAT 1380  
 QY 1381 tgccttaggtcaaggtctgttctactcagagtgatgataataattctagaccggcctatc 1440  
 Db 1381 TGCCTTAGGTCAAAGGTCTGTGTCTACCTCAGAGTGATAATAATCTTAGACCGGCCCTATAC 1440  
 QY 1441 accaaaTaccagggcactagaccccaagtttcaagccaccaataaaTaccgaagcggggtcc 1500  
 Db 1441 ACCAAATACCCAGGCACACTAGACCCCAAGTTTCAAGCCACCACAATAATACGAAGCGGGTCC 1500  
 QY 1501 actttccattggttagtaactctatcaaaogtaaaatcaatcgtcgggtcaggtgttaa 1560  
 Db 1501 ACTTTCCATTGGTAGTAACCTCTATCAAAACGTAATAATCATCAATGTCGGTGCAGGTGTAA 1560  
 QY 1561 taaaaccggtcgggtcaaatgtgcacagctagagcggtgtgaagtggggtcgaagagcg 1620  
 Db 1561 TAAACCGGATGCGGGTCAATGTGCGCAGCTAGAGCGGGTGGTGAAGTGGGCTAAGGAGCG 1620  
 QY 1621 tagaattacttttcaggggtgatataacagctactgcgttaaaataggttttgataatc 1680  
 Db 1621 TAGAATTACTTTTCAGGGTGATGATAACAGTACTGAGCTAAAAATAGGTTTGGATAATAC 1680  
 QY 1681 tttaactattaaaggttggtgcagagccaaacgcatttaacccgataataatcgggtgtgt 1740  
 Db 1681 TTTAACCTATTAAAGGTGGTGCAGAGACCAACGCATTAAACCGATAAATAATATCGGTGTGT 1740  
 QY 1741 aagaaggtgataaattagtggtctgaagttaaactgtctaaacttttaacaattctac 1800  
 Db 1741 AAAAGAGGCTGATAAATAGTGGTCTGAAGTTAAACTTTGCTTAAACCTTTTAAACAATCTTAC 1800  
 QY 1801 tgaggtgaatcacactacattaaatgcacaacacagtttaaggttagttagtagtag 1860  
 Db 1801 TGAGGTGAATACAACCTACATTAAATGCCACACACAGTTTAAAGGTAGGTAGTAGTAG 1860  
 QY 1861 tactacagctgaattatttgtagtgatagtttaacctttaccgccccaatacagggcagtc 1920  
 Db 1861 TACTACAGCTGGAATTATTGAGTGATAGTTTAACTTTTACCCTTACCCTTAAATACAGCGAGTCA 1920  
 QY 1921 aagcacaagcaaaacgctctatggcttaagtgggtgaagtttacttaataatcagaac 1980  
 Db 1921 AACACAAAGCAAAACCGCTCTATGGCGTTAATGGGGTGAAGTTTACTTAATTAATGAGAAAC 1980  
 QY 1981 aacagcagcaatcggcactactcgtattaccagagataaaatggctttgctcagagatg 2040  
 Db 1981 AACAGCAGCAATCGGCACCTACTCGTATTACCAGAGATAAAATTTGGCTTTGCTCGAGATGG 2040  
 QY 2041 tgaattgtgaTgaaaaaaagcaccatatttggataaaaaaacaacttaagtggttagtgt 2100  
 Db 2041 TGATGTTGATGAATAAACAGCACATATTGGGATAAAAAACAACCTTAAAGTGGGTAGTGT 2100  
 QY 2101 tgaattaccatgacaatggcattgatgcaggttaataataaaagatcagtaattctggcaa 2160  
 Db 2101 TGAATTTTACCATAGACAATGGCATTTGATGCAGGTAAATAAAAAAGATCAGTAATCTTGCCTAA 2160  
 QY 2161 aggtagcagtgctaaagcagtcggttaccatcgacagctcaagcgccgaagcctacttt 2220  
 Db 2161 AGGTAGCAGTGTCAACGATCGGTTACCATCGAACAGCTCAAGCGCCCAAGCCCTACTTTT 2220

QY 2221 aaacgcagggcgctggcatcagtgctcacacctactgaaatatcagcttgatgctgaagatgg 2280  
 Db 2221 AAACGCAGGGCGCTGGCATCAGTGTACACCTACTACTGAAATATATCAGTTGATGTAAGAGTGG 2280  
 QY 2281 caatgttaccgcgccccaaacttacaacattggcgtgaaacaccacggcgttcaacagtgatgg 2340  
 Db 2281 CAATGTTACCGCCCCAACTTTACAACATTGGCGTGAACACCACCGAGCTTAAACAGTGATGG 2340  
 QY 2341 cactagtataaaatttagtttaagggttagtgggtacgaaacaataagcttagttaccgcccga 2400  
 Db 2341 CACTAGTGATAAATTTAGTGTTAAGGCTAGTGGTACGAACAATAGCTTAGTTACCGCCGA 2400  
 QY 2401 acatttggcagctatctataatgaagtcacgcagcgctgacagtgctctacaaaagctt 2460  
 Db 2401 ACATTTGGCAAGCTATCTAAATGAAGTCAATCGAACGGCTGACAGTGTCTCTACAAAGCTT 2460  
 QY 2461 tacggtttaaagaagaacacatgatgacgcaacgctctacacgtgcttaagatacagac 2520  
 Db 2461 TACCGTTTAAAGAAGAACGATGATGACGCCAACGCTATCACCGTGGCTAAAGATACGAC 2520  
 QY 2521 aaaaaaTgcgggcgagtcagcatcttaaaactcaaaaggttaaaacgcgtctaaacggttgc 2580  
 Db 2521 AAAAAATGCCGGCGCAGTCAGCATCTTAAACTCAAAGGTAAACACGGTCTAACCGTTGC 2580  
 QY 2581 taccaaaaagatgggttacggttacccttgggcttagccaaagatagcggtctgcaccattgg 2640  
 Db 2581 TACCAAAAAGATGGTACGGTTACCTTTGGGCTTAGCCAAAGATAGCGGTCTGACCATTGG 2640  
 QY 2641 caaaagcaccctaaacacacatggcttgactgtttaaagataccaaacgaacaaatccaaagt 2700  
 Db 2641 CAAAAGCACCCTTAAACAAACGATGGCTTGACGTGTAAAGATACCAACGAACAAATTCCAAGT 2700  
 QY 2701 cgggtcctaattggcattaaattactaatgtgaatggtagtaacccaggtactcgtgcatgg 2760  
 Db 2701 CGGTGCTAATGGCATTAAATTTACTAATGTGAATGGTAGTAATCCAGGTACTGCGATTGC 2760  
 QY 2761 aaatccgcctgcacattaccagagataaaattggcttgcgttgcgtgctgagtgagttga 2820  
 Db 2761 AAATACCCTGCGATTACCAGAGATAAAATTTGGCTTTGCTGTTGTTGTCAGTTGA 2820  
 QY 2821 tacaacaaaccttatctgaatcaagaacagctacaagttggcaattgaagattaccaa 2880  
 Db 2821 TACAACAAACCTTATCTTATCAAGCAAGCTACAAGTTGGCAATGTTAAGATTACCAA 2880  
 QY 2881 cactggcatlaacgcaggtgtgtaagccatcacagggtgtcccccaacactgcctagcat 2940  
 Db 2881 CACTGGCATTAAACGCAGTGTGTAAAGCCATCACAGGGCTGTCCCCACACTGCGCTAGCAT 2940  
 QY 2941 tgcgcatcaaaagtaggcgcgaacatagaaactgggcaatcacatccaagaacaaagaatc 3000  
 Db 2941 TGCGGATCAAAAGTAGCGGCAACATAGAACTGGGCAATACAAATCCAAAGAACAAATC 3000  
 QY 3001 caacgctgccagcattaatgatataataacaggttttaacctaataaaataatacaaa 3060  
 Db 3001 CAACGCTGCCAGCAATTAATGATATATTAAATPACAGGCTTTAAACCTTAAAAATAATAACAA 3060  
 QY 3061 ccccatgactttgtctccactttagcattgttgactttgccaatggcaatggcaccac 3120  
 Db 3061 CCCCATGACTTTGTCTCCACTTATGACATTGTTGACTTTGCCAATGGCAATGCCACCAC 3120  
 QY 3121 cgccagtagtaacccatgataccgctaaacaaacccagtaaagtggtatgatgtgaatgt 3180  
 Db 3121 CGCCACAGTAGTAAACCCATGATACCGCTAAACAAACCCAGTAAAGTGTATATGATGTGAATGT 3180  
 QY 3181 ggaatgatacaacattcatctaacagcactgatgacaataaaacacttggcgtcaaaac 3240  
 Db 3181 GGATGATACAAACCTTATCTTAACAGGCACCTGATGACAATAAAAAAATTTGGCGTCAAAAC 3240  
 QY 3241 caccaaaactgaacaaaaaagtgctaatgtgtaataacagcaactaactttaaagttaactc 3300  
 Db 3241 CACCAAACTGNACAAAAACAGTGTAAATGTTAATACAGCAACTAACTTAAATGTTAACTC 3300  
 QY 3301 tagtgaagaagatgcccttgttaaagcgcgaagacatgcgcgaaaaatctaaacacccctagc 3360

|||||  
Db 3301 TAGTGATGAGATGCCCTTGTAAAGCCAAAGACATGCCGAAATCTAAACACCCTAGC 3360  
Qy 3361 caaggaattacacaccacaaaggacagacagacacgcgcctacaaacctttacagttaa 3420  
Db 3361 CAAGGAAATCACACCCACCAAGGCACAGCAGACCGCCCTACAAACCTTTACCGTTAA 3420  
Qy 3421 aaaggtagatgaaataataatgctgatgacgcacacgcctacccgltgggtcacaagaa 3480  
Db 3421 AAAGGTAGATGAAATATANTGCTGATGACGCCAACGCCATACCGTGGGTCAAAGAA 3480  
Qy 3481 cgcataataatcaagtcaaaccccttaaacactcaaaaggtgaaacaggtcttaataataaac 3540  
Db 3481 CGCAATAATCAAGTCAACACCCCTAAACACTCAAAAGGTGAAACGGCTCTTAATATTAAAC 3540  
Qy 3541 cgcataaagtgcaggttaccttggcttgcattacacacacacagcgtcttaagcgcgcaa 3600  
Db 3541 CGCAATAATGGTAGGTTACCTTTGGGANTTAACACCAAGCGGCTTTAAAGCCGGCAA 3600  
Qy 3601 aagcacctaaacgacggtggtctgtctattaaaaacccccactgtagcgaacaaatcca 3660  
Db 3601 AAGCACCTAAACGACGGTGGCTGTCTATTAAAAACCCCACTGGTAGCGAAATCCA 3660  
Qy 3661 agtcggtgctgagtgaggttgcaggttgcaggttgcaggttgcaggttgcaggttgcag 3720  
Db 3661 AGTCGGTGTGATGGCGTGAAGTTTGGCCAGGTTAATAATATGCTGTGTAGGTGCTGG 3720  
Qy 3721 cattgatgcaactgcattaccagagtgaaatggctttactggactaatggctc 3780  
Db 3721 CATTGATGGCAACACTGCCATTACAGAGATGAATTTGGCTTTACTGGGACTTAATGGCTC 3780  
Qy 3781 acttgataaaagcaaacccccacctaagcaagagcgccattaacgcaggtggttaaaagat 3840  
Db 3781 ACTTGATAAAAGCAAAACCCCACTTAAGCAAAAGCGCATTAACGCGAGTGTAAAGAT 3840  
Qy 3841 tacaacaattcaatcaggtgagattgccccaaacagccatgactgctgacagcgccaa 3900  
Db 3841 TACCACAAATTCAATCAGTGAGATTGCCAAACAGCCATGATGCTGTGACAGCGGGCAA 3900  
Qy 3901 gatttatgattaaaaacgcgaacttgaacacacacacacacacacacacacacacacac 3960  
Db 3901 GATTTATGATTTAAACACCGAACTTGAAACAAATCAGCAGTACTGCCAAACAGCACA 3960  
Qy 3961 aaactcattacacgaattctcagtagcagatgaacaaaggttaataactttacggttagta 4020  
Db 3961 AAACCTCATATACGAATTTCTCAGTAGCAGATGAACAAGGTAAATAACTTTACGGTTAGTAA 4020  
Qy 4021 cccctactccagttatgacacctcaaacacacacacacacacacacacacacacacacac 4080  
Db 4021 CCTTACTCCAGTTATGACACCTCAAGACCTCTGATGTATCATTACCTTTGACGGTGAANA 4080  
Qy 4081 cggcattaccacacaggttaataaagggtggtgctggtggtggtggtggtggtggtggtggt 4140  
Db 4081 CGGCATTACCACCAAGGTAAATAAGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 4140  
Qy 4141 cttaacacgcgttaagctgacgctggtggtggtggtggtggtggtggtggtggtggtggtg 4200  
Db 4141 CTTAACACCGCTTAAGCTGACCGTGGGTGAATAATGATGCAAGAGGCTTGTCTTTGACAG 4200  
Qy 4201 ccaaaatggtcaaataccatcacagagactaagcaaacacttagtcaatggttaccatgta 4260  
Db 4201 CCAAAATGGTCAAAATACCATCAGAGACTAAGCAACACTCTAGCTAATGTTTACCNAATGA 4260  
Qy 4261 taaaggtgagctacgcacacacagagcaggtgcaataataataaagacgaagacaaacccg 4320  
Db 4261 TAAAGGTAGCTAGCCACACAGAACAGGGCAATATAATCAAGACGAAGACAAACCCG 4320  
Qy 4321 tgcgcgaagcattgtgattgctgaagcgcaggtttaaacttgaaagcgaatggtgaagc 4380  
Db 4321 TGCCGCCAGCATTTGTTGATGTGCTTAAGCCCGAGGCTTTAACTTGCAAGCAATGTTGAAGC 4380  
Qy 4381 ggttgacttgcctcacttatgacacgcgtcaacttggcagtggaatgacacacgcgc 4440  
|||||

Db 4381 GGTGACTTTGTCTCCACTTATGACACCGTCAACTTTGCCGATGGCAATGCCACACCGC 4440  
Qy 4441 taagtgacctatgatgac 4500  
Db 4441 TAAAGTACCTTATGATGAC 4500  
Qy 4501 tgatataccacttgaagttaaaagataaaaaaacttggcgtaaaaaacacacacacacacac 4560  
Db 4501 TGATACAACCAATTCAAGTTAAAGATAAAAAAATTGGCGCTTAAACACACACACATTTGACCAG 4560  
Qy 4561 tactggcacaggtgctactaataatttgccttaagcaatacaagctactggcgatgagcttgt 4620  
Db 4561 TACTGGCACAGGTGCTAATAAATTTGCCCTAAGCAATCAAGCTACTTGGCGATGGCGTTGT 4620  
Qy 4621 caagccagtgatgctgctcactcaaaacacacacacacacacacacacacacacacacac 4680  
Db 4621 CAAGGCCAGTGTATGCTGTCTCATCTAAACACCTTATCTGGGACATCCAAACTGCCAA 4680  
Qy 4681 aggggcaagccgaacgaactcagcaggtgtatgtggtgctgctgctgctgctgctgctgct 4740  
Db 4681 AGGGCAAGCCGAAGCAACAACTCAGCAGGCTATGTGGATGCTGATGGCAATTAAGGTCTAT 4740  
Qy 4741 ctatgacagtaccgatacaagaactactatcaagccaaaatgatggcacagttgataaac 4800  
Db 4741 CTATGACAGTACCGGATAACAAGTACTATCAAGCCAAAAAATGATGGCACAGTTGATAAAC 4800  
Qy 4801 caaagaagttccaaagacaaactggtgcgcaagcccaaacccacagatggcacattggc 4860  
Db 4801 CAAGAAGTTGCCAAAGACAACTGGTGCCTCAAGCCCAACCCACAGATGGCACATTGGC 4860  
Qy 4861 tcaaatgaatgtcaaatcagtcattcaacaaagaacaaatgaatgatgctgctgctgctgctg 4920  
Db 4861 TCAATGAATGTCAAAATCAGTCATTACAAAGAAACAAGTAAATGATGTCATTAATAAGCA 4920  
Qy 4921 aggcataataaagacacacgcttgtttaaaggacttgaagacacacacacacacacacacac 4980  
Db 4921 AGGCATCAATGAAGACACAGCGCTTTGTTAAAGGACTTGAAGAGCGGCTTCTGTGATAACA 4980  
Qy 4981 aacaaacacgcgcagtaactggtggtggtggtggtggtggtggtggtggtggtggtggtg 5040  
Db 4981 AACCAAAACCGCCAGTAACCTGTGGGTGATTTAAATGCCGTTGCCCAACACCCCTGAC 5040  
Qy 5041 ctttgcagggatgacagac 5100  
Db 5041 CTTTGCAGGGATACAGCACACAGCGTTAAAAAATGGCGGAGACTTTTACCATCAAAAGG 5100  
Qy 5101 tgggcaaac 5160  
Db 5101 TGGGCAACACACACACACATAGCTAACCGATTAATTAACATCGGTGGTGGTAGCAGGTACTGA 5160  
Qy 5161 tggcttcaactgtcaaaacttgcac 5220  
Db 5161 TGGCTTCACTGTCAAACTTGCCAAAGACTAACCAATCTTAACAGCGTTAATGTCAGGTGG 5220  
Qy 5221 caccac 5280  
Db 5221 CACCAAAATGATGACAAAGCGGTGCTTTTGTAGACTCAAGCGGTCAAGCCCAAGCAAA 5280  
Qy 5281 caccctgtgcttaagtgcac 5340  
Db 5281 CACCCCTGTCTAGTGCCATGGCTGGACCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 5340  
Qy 5341 caaaggcacaagaataccgacgctgccaatgtcaacacacacacacacacacacacacacacac 5400  
Db 5341 CAAAGGCCAAAGATACCGCGCTGCCAATGTACAACAGTTTAAACGAAGTACGCAACTT 5400  
Qy 5401 gttgggtcttgtaatgctgtaataataacacacacacacacacacacacacacacacacacac 5460  
Db 5401 GTTGGGTCTTGGTAATGCTGGTAATGAACCTGACCGCAATACAGGTAAACATTTGCCGA 5460  
Qy 5461 catcaaaaaagac 5520  
Db 5461 CATCAAAAAGACCCCAATTCAGGTTTCATCTTAACCGCACTGTCTATCAAAAGAGGCAC 5520



Qy	6601	tggtcaatgggtatttaaatacaatggttcacgccgataccccaggccatgtagggggcgc	6666
Db	6601	TGgTCAATGGgTATTAAATCAATGGTTTCAGCCGATACCCAAAGGCATGTAGGGGGCGC	6666
Qy	6661	agttggtcaggttttcacttttaagccataaatccaaagattttactttaaatacaatc	6720
Db	6661	AGtTGTGcAGGTTTCACTTTTAAGCCATAATCCAAAGATTTTACTTTAAATCAATC	6720
Qy	6721	tcaccatagttgtataaaacagcatcagcatcagtcattactcttcaacatgacccaaatcgccattg	6780
Db	6721	TCACCATAGTTGTATAAAACAGCATCAGCATCAGTCATATTACTGATGCTGATGTTT	6780
Qy	6781	atcaactaaaccattttacccgtcgaagtattctcttcaacatgacccaaatcgccattg	6840
Db	6781	ATCACTTAAACCATTTTACCGCTCAAGTGATCTCTTTCAACATGACCAATCGCCATTG	6840
Qy	6841	atcataggtaaacttattgataaattttatacaatagtttgttagatatggtttaaatt	6900
Db	6841	ATCATAGGTAAACTTTATTGAGTAATTTTATCAATGTAGTTGTAGATATGGTTAAAT	6900
Qy	6901	gtgccattgacccaaaaatgaccgatttatcccgaaaaattctgattatgatacgttgac	6960
Db	6901	GTGCCATTGACCAAAAAATGACCGATTATCCGAAAAATTTCTGATTATGATCCGTTGAC	6960
Qy	6961	ctgcagggtcgac	6972
Db	6961	CTGCAGGTGAC	6972
RESULT	2		
LOCUS	AR040716	6973 bp	DNA
DEFINITION	Sequence 1 from patent US 5808024.		PAT
ACCESSION	AR040716		
VERSION	AR040716.1	GI:5960079	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 6973)		
AUTHORS	Sasaki,K., Harkness,R.E., Loosmore,S.M. and Klein,M.H.		
TITLE	Nucleic acids encoding high molecular weight major outer membra		
JOURNAL	protein of moraxella		
FEATURES	Patent: US 5808024-A 1 15-SEP-1998;		
	Location/Qualifiers		
source	1..6973		
BASE COUNT	2265 a 1555 c 1533 g 1620 t		
ORIGIN	/organism="unknown"		
Query Match	99.8%;	Score 6961;	DB 9; Length 6973;
Best Local Similarity	100.0%;	Pred. No. 0;	
Matches 6972;	Conservative 0;	Mismatches 0;	Indels 1; Gaps
Qy	1	ccatgatatggcaggtgtgctgcgcctgcgtatgatggcgtacacccccatttgcgc	60
Db	1	CCATGGATATGGCAGGTGTGCTCGCCTGCCTGATGATGCGATGACACCCCATTTGCC	60
Qy	61	catatctgtacgatttgacatgtgatattttaaactgtgacatgattttaaactgtttt	120
Db	61	CATATCTGTACGATTTGACATGTGATGATGATTAAACATGTGACATGATTAAACATTC	120
Qy	121	aatactgttgccatcattaccataaatttagtaacgcatttagtaacgcatttgaataat	180
Db	121	AATACGTGTGCCATCATTTACCATTAATTTAGTAACGCATTTAGTAACGCATTTGTAA	180
Qy	181	cattgcgcctttagttatcatatgaatgaatattatgattgattctgattattgt	240
Db	181	CATTGCGCCCTTATGTGTATCATATGAATAATAATATGATTGTATCTGATTATTGT	240
Qy	241	atcagaatgggtgatgcataatgatgcctacgattgattgggttaactactctatg	300

Db 241 ATCAGAAATGGTGATGCTATATATGATGATGCCCTACGAGTGTGATTTGGGTTAATCACTCTATG 300  
Qy 301 atttgatataatttgaaactaatctattgactttaataacocatatggtttataatttagca 360  
Db 301 ATTTCGATATATTTTGAAACTAATCTATTTGACTTAAATCACCATATGTTATATATTTAGCA 360  
Qy 361 taatgtaggtgttttttgaaataatcatatgcatatgcttctactgttactaccatgct 420  
Db 361 TAATGGTAGGCTTTTGTGTAATAATCACAATGCAATATTTGTTCTACTGTACTACCATGCT 420  
Qy 421 tgaatgacgatcccaatcaccagatttcaattcaagtgatgtgtttgtatagcaccattta 480  
Db 421 TGAATGACGATCCCAATCACCAGATTCAATTCAGTGATGTGTTTGTATACGCACCAITTA 480  
Qy 481 ccctaattatttcaatcaaatgccttatgtcagcatgtatcatctttttttaaggttaaacac 540  
Db 481 CCTTAATTAATTCATCAAAATGCCATATGTCAGCATGTATCATATTTTAAAGGTAAACCAC 540  
Qy 541 catgaatcatctatataaagtcattttaaacaagccacagccacatttatggcagtgcc 600  
Db 541 CATGAATCACAATCTATAAGTTCATCTTTAACAAAGCCACAGGCACATTTATGGCAGTGCC 600  
Qy 601 aagtagcgcgaatccacagcac-ggggggggtagctgtgtctacagggcaagtggca 659  
Db 601 AGAGTACGCCAATCCACAGCAGCGGGGGGGGTAGCTGTCTACAGGCAAGTTGCA 660  
Qy 660 gtgtatgactctgtgagctttgcccgtattgcccgtctgcgtctgcgtctgcgtgca 719  
Db 661 GTGTATGCATCTGTAGCTTTGCGCGTATTTGCGCGCTCGCTCTCTCGTGTATGCGTGCA 720  
Qy 720 cgtcagtgagcagctgtctcctaaataaagatacacaacatatcgcaattgtggaac 779  
Db 721 CGCTAGTGGCAGTGTCTATGTCTCAAAAAAAGATACCAAAATATCGCAATTTGGTGAAC 780  
Qy 780 aaacagcagcaagcgtcagcactgcgaagcggaagcggtgatgcagccattgctattg 839  
Db 781 AAAACAGCCAGACGCTCAGGCATTCGCAAGCGGACGGTATCGAGCCATTTGCTATTG 840  
Qy 840 gtgaataatgctaacgacagcggtgcagccatcgccatcggttagtagtaataaaactg 899  
Db 841 GTGAAATGTCTAACGCACAGCGGCGGTCAAGCCATCCCATCGGTAGTACTAATAAACTG 900  
Qy 900 tcaatgaagcagctttggtaagataggttacogattgctacgggtcaagagtcaccatgcca 959  
Db 901 TCAATGGAAGCAGTTTGCATAAGATAGGTACCGATGCTACGGGTCAAGAGTCCATCGCCA 960  
Qy 960 tcggtgtgtatgaagctagtggtgattcctcgattgcatcggttagtgacttac 1019  
Db 961 TCGGTGTGTATGTAAGGCTAGTGGTGTGTCCTCGATTGCCATCGGTAGTGACTTAC 1020  
Qy 1020 atttgcttgatcagcatggtaacttaacatccgaaggtactctgtattacagcatctta 1079  
Db 1021 ATTTTGCTTGATCAGATGGTTAATCCTAAATCCGAATCCGAAGGTACTCTGATTTAAGCATTTA 1080  
Qy 1080 ttaacggccatgcatattaaagaaatacgaagctcaaggaataatgatgtaaaatata 1139  
Db 1081 TTAACGGCCATGCGAGTATTAAGAAATAATACGAAGTCAAAGGATAATGATCTAAATATA 1140  
Qy 1140 gacgcacaacccgcaagcgacacgcccagtcagctggggagccatgtcatatgcacag 1199  
Db 1141 GACGCACAACCGCAAGCGGACACGCCAGTACTGTCAGTGGGAGGCCATGTCATATGCACAG 1200  
Qy 1200 gtcatatttccaacgctttgtgacagggcaacagctaaagtcctattctcttgagcag 1259  
Db 1201 GTCAATTTTCCAAACGCCCTTGTACACGGGCAACAGCTTAAAGTGGCTATTCTCTTGGCAG 1260  
Qy 1260 lgggcttgcgcacacagcgagggcaatctacaactcgctattggttctgtatgcaacat 1319  
Db 1261 TGGGTCTTGGCCGCACACCGGAGCCCAATCTACAATCGCTATTGTTGTTCTGATGCAACAT 1320  
Qy 1320 ctatgctgtgggagcgtatgccccttggtgaggtactcgtcgtcagctacagggagta 1379  
Db 1321 CTAGCTCGTTGGAGCGATAGCCCTTGTGTGAGGTACTCTGCTCAGCTACAGGGCAGTA 1380

Qy 1380 ttgccctaggtcgaagttctgttgcactcagagtgaataaattcttagaccgacctata 1439  
Db 1381 TTGCCCTAGGTCAAGTTCTGTGTCTACTCAGAGTGATAATAATAATCTAGACCGGCTATA 1440  
Qy 1440 cacaataatccaggcactagaccaccaagtttcaagccaccaataataacgaagcggggc 1499  
Db 1441 CACCAAAATCCAGGCACGTAGACCCCAAGTTTCAGCCACCAATAATAGCAAGGGGGTC 1500  
Qy 1500 caatttccattgttagtaactctatcaaacgttaaacatcctcaatgctcggtgcaggtgtta 1559  
Db 1501 CACTTTCCATTTGTGTAGTAACCTCTATCAACGTAATAATCATCAATGTCGCTGCAGGTGTTA 1560  
Qy 1560 ataaacccgagtcggttcaatgtggcacagctagaacggtggtggtggtggtggtggtggt 1619  
Db 1561 ATAAACCGATGCGTCAATGTGGCAGCTAGAACGGTGTGTGAAGTGGCTTAAGGAGC 1620  
Qy 1620 gtgaattacttttcagggttgatgaataacagtaactgaactgaataaagtttggataata 1679  
Db 1621 GTAGAAATTAATTTTCAGGGTGATGAACAGTACTGACGTAAAAAATAGGTTTGGTANTATA 1680  
Qy 1680 ctttaactattaaaggtggtgcagagaccacacgatttaaccgataataatcgtgtgtg 1739  
Db 1681 CTTTAACTATTAAAGTGTGGTGACAGAGACCAACGATTAAACGATATAATAATATCGGTGTG 1740  
Qy 1740 taaaagggctgataatagtggtgtgaaagtctaaacttgctaaacatttaacaacattta 1799  
Db 1741 TAAAGAGGGCTGATAATAGTGGTCTGAAAGTTAAACTTGTCTAAACCTTTAAACAATCTTA 1800  
Qy 1800 ctgaggtgaataacacactacattaaatcccaacacacagtttaaggttaggtagtagta 1859  
Db 1801 CTGAGGTGAATACAACTACATTAANTGCCACACACACAGTTAAGTATAGTAGTAGTA 1860  
Qy 1860 gtactcagctgaattattgagtgatagtttaaccccttaccagcccaatacagggcagtc 1919  
Db 1861 GFACTACAGCTGAATATTGAGTGTATAGTTTAACTTTTACCAGCCCAATACAGGCGATC 1920  
Qy 1920 aaagcacaagcaaaacgctctatgctgaatgggtgaaagtcttaataatgcagaaa 1979  
Db 1921 AAAGCACAAAGCAAAACCGCTCTATGGCGTTAATGGGTGAACTTTACTAATAATACACAAA 1980  
Qy 1980 caacagcagcaatcgccactactcgattaccagagataaaatggcttgcagagtg 2039  
Db 1981 CAACAGCAGCAATCGGCATCTACTCGTATTACAGAGATATAAATTTGGCTTTGCTCGAGATG 2040  
Qy 2040 gtgactgtgataaaacaacagccatatttggtataaaacaacttaaaagtggtgagtg 2099  
Db 2041 GTGATGTTGATGAAACCAAGCCCTATTATTGGATAAAAACAACTTAAAGTGGGTAGTG 2100  
Qy 2100 ttgcaattaccatagacaaatggcattgtatgcagggttaaataaaagatcagtaacttgcca 2159  
Db 2101 TTTGCAATTAACCATACACAATGGCATTTGATGAGGTAAATAAAAGATCAGTAATCTTGCCA 2160  
Qy 2160 aaggtagcagtgctaacgatgcggttacatcgaaacagctcaaaagccgacgactactt 2219  
Db 2161 AAGGTAGCAGTGTCTAACGATGCGGTTTACCATCGAACAGCTCAAAAGCCCAAGCCCTACTT 2220  
Qy 2220 taaacgagcgctggcatcaggtcacaccctactgaaatatacagttgatgtaagagtg 2279  
Db 2221 TAAACGCGAGCGCTGGCATCAGTGTGCACACCTACTTGAATAATATCAGTTGATGCTAAGAGTG 2280  
Qy 2280 gcaatgttaccgcccccaacttacaacattggcggtgaaacacccagcttaacagtgatg 2339  
Db 2281 GCAATGTTTACCGCCCCCACTTACAACATTTGGGCTGAAAAACCCAGCGCTTAAACAGTGATG 2340  
Qy 2340 gcaatgtgataaaatttagtttaagggtagtgggtacgaacaataatagcttagtaccgag 2399  
Db 2341 GCACTAGTGTAAATTTAGTTTAAAGGTAGTGGTACGAACAATAATAGCTTAGTTACGCGCG 2400  
Qy 2400 aacatttggcgaagctatctaaatgaaggtcaatcgaaacggtgacagtgctctacaaagct 2459  
Db 2401 AACATTTGGCAAGCTATCTAAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 2460

QY 2460 ttaccggttaaagaagaagcagatgatgacgccaacgctatcacccgtggtctaaagatacaga 2519  
DB 2461 TTACCCTTAAAGAAGAAGACGATGATCAGCCCAACGCTATCACCGTGGCTAAAGATACGA 2520  
QY 2520 caaaatgccggtcagtcagcatctaaactcaaaagtaaaaaacgctctaaacggttg 2579  
DB 2521 CAAAATGCCGCGCAGTCAGCATCTTAAACTCAAAAGTAANAACGGTCTAACGGTTG 2580  
QY 2580 ctacaaaaaagatggttacggtttacgtttgggtttagccaagatagcgtctgaccatttg 2639  
DB 2581 CTACCAAAAAAGATGTACGGTTTACCTTTGGGCTTAGCCAAGATAGCGGTCTGACCATTG 2640  
QY 2640 gcaaaagcaccctaaacaacagatggttgaactgttaaagataccaaacgcaaaatccaag 2699  
DB 2641 GCAAAAGCACCTTAAACAACAGATGGGTGACTGTTAAAGATATACCAACGAACAATCCAAG 2700  
QY 2700 tcggtgctaagtcattaaattactaatgtgaatgtagtaattcaggtcaggtactggcattg 2759  
DB 2701 TCGGTGCTAATGGCATTAAATTTACTAATGTGAATGGTAGTAATCCAGGTACTGGCATTG 2760  
QY 2760 caaataccgctgcattaccagagataaaattggctttgctggttctgtggttcgagttg 2819  
DB 2761 CAAATACCGCTCGCATTTACCAGAGATAAAATTTGGCTTTGCTGTTCTGATGGTGCAGTTG 2820  
QY 2820 atacaacaaaccttatcttgatcaagacaagctacaagttggcaatgttaagattacca 2879  
DB 2821 ATACAACAACCTTATCTTGATCAAGACAAAGCTAGCAAGTTGGCAATGTTAAGATTACCA 2880  
QY 2880 acactggcattaacgcaggtgggtaagccatcacaggtctgtcccaacactccttagca 2939  
DB 2881 AACTGGCATTAAACGAGGTGGTAAAGCCATCACAGGGCTGTCCCAACACTGCTTAGCA 2940  
QY 2940 ttgccgatcaaatagtcgcaacataagaactgggcaatacaatccaaagacaagaacaaat 2999  
DB 2941 TTGCCGATCAAACTAGCGCAACATAGAACTGGGCAATAACAATCCAAAGACAAAGACAAAT 3000  
QY 3000 ccaacgtgccagcattaatgatataataacaggtcttaacctaaacaaataataaaca 3059  
DB 3001 CCAACGCTGCCAGCATTAATGATATATTAATACAGGCTTTAACTTAAACCTTAAACAAATTAATAACA 3060  
QY 3060 acccttgacttgtctccactatgacattgttgacttgcgaatggcaatggcaatggcacc 3119  
DB 3061 ACCCATTTGCTTGTCTCCACTTATGACATTTGTTGACTTTGGCAATGGCAATGCCACCA 3120  
QY 3120 ccgccagtaaacccatgataccgcttaacaaaacccgtaaaagtgggtatgatgtgaatg 3179  
DB 3121 CCGCCACAGTAACCCATGATACCGCTAACCAAAACAGTAAGTGGTATATGATGTGAATG 3180  
QY 3180 tggatatacaaacattcatcttaacagggcaactgatgacataaaaaacttggcgtaaaaa 3239  
DB 3181 TGGATATACAACCATTTCACTAACAGGCACCTGATGACATAAAAAAATTTGGCGTCAAAA 3240  
QY 3240 ccacaaactgaacaaacaaagtgctaatgtgtaatacagcaactaaacttaattgttaact 3299  
DB 3241 CCACCAACTGAACAAACAAAGTGTAAATGGTAAATACAGCACTAACTTTAAATGTTAACT 3300  
QY 3300 ctagtgtgaagtgccttctttaaagcgaagacatcgccgaaaaatctaaacacccctag 3359  
DB 3301 CTAGTCATGAAGATGCCCTTTGTTAAACGCCAAGACATCGCGCAAAATCTAAACACCCCTAG 3360  
QY 3360 ccaagaaaaattcacaccaccaaaggaacagcagacacccgcccctacaaaacctttaccgta 3419  
DB 3361 CCAAGGAAATTTCAACACCACCAAGGCACAGACACACCGCCCTTACAAACCTTTTACCGTTA 3420  
QY 3420 aagagttagatgaaataataatgctgtagcgcgaacccatcacggtgggtcaaaaga 3479  
DB 3421 AAAAGGTAGATGAATAATAATGCTGATACGCCCAACGCCATCACCGTGGGTCAAAAGA 3480  
QY 3480 acgcaaatatcaagtcaaacacccctaaactcaaaaggtgaaaaaggtctttaaattaaaaa 3539  
DB 3481 ACGCAAAATATCAAGTCAACACCTTAACACTCAAGGTGAACGGTCTTAAATATTAATA 3540  
QY 3540 ccgacaaaaatggtacggtttacgttttggcattataacaccacaagcgggtctttaaagccggca 3599

DB 3541 CCACAAAAATGTAGCGTTACCTTTTGGCATTAAACCAACGCGGTCTTAAAGCCGGCA 3600  
QY 3600 aaagcaccctaaacgacggtggctgtctattaaaaacccactggttagcgaacaaatcc 3659  
DB 3601 AAAGCACCTTAAACGACGGTGGCTTGTCTATTAAAAACCCACTGTGTAGGACGAATAATCC 3660  
QY 3660 aagtcggtgctgtagcgtgaaagtttgcgaaggttaataataatggtgtgtgtaggtgctg 3719  
DB 3661 AAGTCGCTGCTGATGGCGTGAAGTTTGCCTCAAGGTTAATAATAATGTTGTTGTAGTGCTG 3720  
QY 3720 gcaattatggcacaactcgcattaccagagatgaaattggctttacttggacttaaggtct 3779  
DB 3721 GCATTGATGCACAACTCGCATTTACCAGAGATGAAATTTGGCTTTACTGGGACTAATGGCT 3780  
QY 3780 cacttataaaagcaaaccccccttaagcaagacagcattaaacgaggtggttaaaaaga 3839  
DB 3781 CACTTGTATAAAGCAAAACCCCTTAGCAAAAGACGGCATTTACGCAAGGTGGTAAAAAGA 3840  
QY 3840 ttaccaacattcaatcaggtgagattgccccaaaacagccatgatgctgtgacagggcgca 3899  
DB 3841 TTACCAACATTTCAATCAGGTGAGATTGCCCCAAAACAGCCATGATGCTGTGACAGCGGCA 3900  
QY 3900 agatttatgatttaaaaaacccgaacttgaacaaacaaatcagcagtagtactgccaacacgac 3959  
DB 3901 AGATTTATGATTTAAAAACCGAATTTGAAAACAAAATCAGCAGTACTGCCAAAACAGCAC 3960  
QY 3960 aaaaactattcacgaattctcagtagcagatgaacaaagtgtaataactttacggttagta 4019  
DB 3961 AAACCTATTTACAGCAATTTCTCAGTAGCAGATGACAAAGGTAAATTAATTTACGGTTAGTA 4020  
QY 4020 acccttactcagttatgacacctcaaaagaccttgatgctcatcacctttgaggtgaaa 4079  
DB 4021 ACCCTTACTCCAGTTATGACACCTCAAAAGACCTCTGATGTCATCACCTTTGCAGTGCAA 4080  
QY 4080 acggcattacccaacgaagtaaaataaagggtgtgtggtggtggcgcatlgacaaaacaaag 4139  
DB 4081 ACGGCATTTACACCAAGGTAAATAAAGGTGTGTGCGTGTGGGCATTGACCAAAACAAAG 4140  
QY 4140 gcttaacccagcctaaagctgacgtgggttaataatggcgaaggtcattgtcattgaca 4199  
DB 4141 GCTTAACCAAGCCTTAAGTCACCGTGGGTAAATATATATGCAAGGCAATTTGTCTATTGACA 4200  
QY 4200 gccaaaaatgttcaaaaatcaccatcacaggaactgaagcaacacttagctaaatgttaccatg 4259  
DB 4201 GCCAAAAATGCTCAAAATACCATCACAGACTTAAGCAACACTCTAGCTAATGTTACCAATG 4260  
QY 4260 ataaaggtagcgtacgcaccacagaacagggcaataataatcaaaagacgaagacaaaaacc 4319  
DB 4261 ATAAAGGTAGCGTAGCGCACACAGAACAGGGCAATATAATCAAAAGACGAAGACAAAACCC 4320  
QY 4320 gtcccgccagcattgttgatgtgctaaagcgcaggtttaaacttgcgaatggcgaatgggtgaag 4379  
DB 4321 GTCCGCCACCATTTGTGATGTGCTAAGCGCAGGCTTTAACTTGCAGAGGCNAATGGTGAAG 4380  
QY 4380 cgggtgactttgtctccacttatgacaccctcaacttgcgcatggcgaatggcaccaccg 4439  
DB 4381 CGGTTGACTTTGTCTCCACTTATGACACCGCTCAACTTTGCGGATGGCAATGCCACACCG 4440  
QY 4440 ctaaggtgacctatgatgacacaagcaaaacccagtaaaagtggctctatgatgtaagtgg 4499  
DB 4441 CTAAGGTGACTTATGATGACACAAGCAAAACCCAGTAAAGTGGTCTATGATGTCAATGTGG 4500  
QY 4500 atgatacaaacattgaagttaaagataaaaaacttgcgttaaaaaaccacacattgacca 4559  
DB 4501 ATGATACACCATTTGAAGTTAAAGATAAAAACTTGGCGTAAAAAACCCACCATTTGACCA 4560  
QY 4560 gtactggcacaggtgcttaataaaatttcccttaagcaatcaagctactggcgatgcgcttg 4619  
DB 4561 GTACTGCGACAGGTGCTTAATAAATTTGCCCTTAAGCAATCAAGCTTACTGGCGATGCGCTTG 4620  
QY 4620 tcaagggcagtgatatcgttgcctatcaaacaccttatctgtggcagatccaaactgcca 4679

Db	4621	TCAAGGCCAGTGATATCGTTGCTCATCTAAACACCTTATCTGGGACATCCAAATGCCA	4680	5760	tccatgtcaacgatgcaatcaagcctgtgtgtacaagggcgtaacggcattgaactcaa	5819
Qy	4680	aaggggcaagcgaacgaactcagcaggtctatgttggtgtgtatggaataaagttca	4739	5761	TCCATGTCAACGATGGCAATCAAGAGCCTGTGTGAAGGGCGTAACGCGATTGACTCAA	5820
Db	4681	AAGGGCAAGCCAAAGCAACACTCAGCAGGCTATGTGGATGTGTGGAATAAGGTCA	4740	5820	gtgctcaggaagcactcagtggtgagtaggtttccagggccaagcgagatggtggaagcgg	5879
Qy	4740	tctatgacagtaaccgataaagaagtaactatcaagccaaataatgatggcacagttgataaaa	4799	5821	GTGCCC'TCAGCAAGCAACTCAGTGGGGATAGGTTTTCAGGCCAAGCGCAGATGCTGAAGCCG	5880
Db	4741	TCTATGACAGTAGCCGATAAACAGTACTATCAAGCCAAATATGATGGCACAGTTGATAAAA	4800	5880	ccgttgaccatagcgagacaaaaccagcaggcaaccaa'tccatcgccatcggtgtaaacg	5939
Qy	4800	ccaaagaagtTgcgaagacaaaactggTcgcccaagcccaaaccccgatggcaattgg	4859	5881	CCGTTGCCATAGGCACACAAACCCAAAGCAGCAACAA'TCCATCGCCATCGGTGATAACG	5940
Db	4801	CCAAACAGATTGCCAAAGACAAACTGGTGCGCCAAAGCCCAAAACCCAGATGGCACATTGG	4860	5940	cacaagccacggggcgatcaatccatcgccatcgggtacaggaatgtggtagcagatgaag	5999
Qy	4860	ctcaaatgaatgtcnaatcagtcattacaagaagaacaaagtaaaTgaTgccaataaaagc	4919	5941	CACAAGCCACGGGCGATCAATCCATCGCATCGGTACAGGCAATGTGGTAGCAGGTAGC	6000
Db	4861	CTCAAATGAATGTCAAAATCAGTATTAAACAAGAACAAAGTAAATGATGCCAATAAANAGC	4920	6000	actcgtgtccatcgcgacccaagcactgttaaggtctgatacaacagttacagtggtgga	6059
Qy	4920	aagcatcaatgaagacacagcccttTgttaaggacttgaaaaagccgcttctctgataaca	4979	6001	ACTGTGGTGGCATCGCGACCCAAAGCACACTGTTAAGGCTGATAACAGTTACAGTGTGGTA	6060
Db	4921	AAGGCATCAATGAAGACAAACGCCTTTGTAAAGGACTTGAAGAGCGCTTCTGTATAACA	4980	6060	ataacacacagtttacagatgccaactcaaacccagatgtctttggTgtgggcaataaacatca	6119
Qy	4980	aaacccaaacccgcagtaactgttggttgatattaaatgcgcgtTgccccaaacacgcgtga	5039	6061	ATAAACACCAAGTTTACCGATGCCACTCAAAACCGATGTCTTTGGTGTGGGCAATAACATCA	6120
Db	4981	AAACCAAAACCCGCAGTAACGTGTGGTGTATTTAAATGCCGTGTGCCCAAAACCCGTGA	5040	6120	ccgtgaccgaagaactcgtgttccttaggttcaaaactctgcatacagTgcagggcacac	6179
Qy	5040	cccttgcaagggatcacaggcacacagcgtcaaaactggcgagactttgacatcaaacg	5099	6121	CCGTGACCCGAAGTAACTCGGTTGCCTTAGGTTCAAACTCTGCCATCAGTGCAGGCACAC	6180
Db	5041	CTTTTGACGGGATACGGCACACACGGCTAAANAACCTGGCGAGACTTTGACCATCAAG	5100	6180	agcaggcacacaaagccaaataatctgacggcgacagcaggTacaacccacacagcagggtg	6239
Qy	5100	gtgggcaaacagacacaaataagctaaacgataataacatacgtgtgtgtgtagcaggtactg	5159	6181	AGGCAGGCACACAAAGCCCAAAATAATCTCACGGCACAGCAGGTACAAACCCACACAGAGGTG	6240
Db	5101	GTGGGCAACAGACACCAATAAGCTAAACGGATAATAACATCGTGTGTAGCAGTACTG	5160	6240	caacccgttaggttaaaagccttTgctgagcaaacggcggtTgtgtgaggtctccgtgggtg	6299
Qy	5160	atggtctactgtcaaatTgcgaagacctaaccatcttaacagcgttaagtgcaggtg	5219	6241	CAACCGGTACCGGTTAAAGGCTTTCTCTGCACAAACGGCGGTTGGTGCCTCTCCGTGGGTG	6300
Db	5161	ATGGCTTCACTGTCAAACTTGGCCAAAGACTTAACCAATCTTAACAGCGTTAATGCAGGTG	5220	6300	cctcaggtTgctgaaacgcgtatccaaaatgtggcagcaggtTgaggttcagTgcacacagca	6359
Qy	5220	gcacccaaatTgtatgcaaaagcgTgctctttgttagactcaagcgggtcaagccaaagcaa	5279	6301	CCTCAGGTGCTGAACGCGGTATCCAAATAATGTGGCAGCAGGTGAGGTGAGTGCACACAGCA	6360
Db	5221	GCACCAAAATGTATGACAAAGGCGTGTCTTTGTAGACTCAAGCGGTCGAAGCAAGCAA	5280	6360	ccgatTcggtgtaaatTggtagccagttgtacaagccacccaaagcatTgcacacgcaacca	6419
Qy	5280	acacccctgtctaatTggtgcgaactggtgtggaagtcacagtcacagtaagtgtg	5339	6361	CCGATGCGCGTCAATGGTAGCCAGTGTGTACAAAGCCACCCAAAGCATTTGCCAAGCCACCA	6420
Db	5281	ACACCCCTGTGCTTAAGTGCCAAATGGGCTGGACCTGGGTGGCAAGGTATCATGTAATGG	5340	6420	atgagcttgacacacTcgtatccacccaaacgaaaaataagggccaaTgcaggggatttcaTcag	6479
Qy	5340	gcaaaagccacaaagataacccagcgtgccaaTgtacaacagtttaaacaggaagTaccgaact	5399	6421	ATGAGCTTGACCATCGTATCCACCAAAACGAAATTAAGGCCAATGCAGGCAATTCATCAG	6480
Db	5341	GCAAAAGCCACAAAGATACCGACGCTGCCAATGTACAACAGTTTAAACGAAGTACGCAACT	5400	6480	cgaTggcgatTgggtccatgccaagcctacattcctctggcagatccatggtTaccggggg	6539
Qy	5400	TgtTgggtctTgttaatTgctgtgtaataagcgtgacggcaatcaggtaaacattgcgg	5459	6481	CGATGGCGATGGCGTCCATGCCAAGCCTACATTTCTTGGCAGATCCATGGTTACCGGGG	6540
Db	5401	TGTTGGGTCTTGTTAATGCTTGGTAATGAATAACGCTGACGGCAATCAGGTAAACATTCGCG	5460	6540	gtattgccaccacacaaacggTcaaggtgcggTggcagtgggactgtcgaagcgtctcggaata	6599
Qy	5460	acatcaaaaagaccacaaatcaggttcacatcTcaaccgcagtcgtcatcaaaagcaggca	5519	6541	GTATTGCCACCCACCAACCGTCAAGGTGCGGTGGCAGTGGGACTGTTCGAAGCTGTCGGATA	6600
Db	5461	ACATCAAAAAGACCCCAAAATTCAGTTTCATCATCTAACGSCACTGTATCAAAAGCAGGCA	5520	6600	atggttcaatTgggtatttaaaatcaatTggttcagccgaTaccocaaagccatgtagggggcg	6659
Qy	5520	cggTactTggcggTaaaggTaaTaaacgataaccgaaaaactTgcactTggtTggtatacaag	5579	6601	ATGGTCAATGGGTATTTAAATCAATGGTTACGCCGATACCCCAAGGCCATGTAGGGGGCGG	6660
Db	5521	CGGTACTTGGCGGTAAAGGTAAATAACGATACCCGAAAAACTTGGCAC'TGGTGGTATACAAG	5580	6660	cagTtggTgagcagtttTcaacttttaagccataaaatcgcaagattttactttaaataatcaat	6719
Qy	5580	TgggcgtgataaagacggcaaacctaacgcgctatttaagcaatgtttTgggtcaaaaccc	5639	6661	CAGTTGGTGACAGGTTTTCATCTTTTAAGCCATAATTCGCAAGATTTTACTTAAATAATCAAT	6720
Db	5581	TGGGCGTGGATTAAGACGGCAACCGCTAACGGCGATTTAAGCAATGTTTGGGTCAAAACCC	5640	6720	ctcaccaatgtTgtataaaacagcatcagcaTcagTcaTattactTgaTgctgatttttt	6779
Qy	5640	aaaaagatggcagcaaaaaagccctgtctgcacactataacgcgcgaggtTcaagcaact	5699	6721	CTCACCATAGTTGTATAAAAACAGCATCAGTCATATATTACTGATGCTGATGTTTTT	6780
Db	5641	AAAAGATGGCAGCAAAAAGCCCTGTCTGCCACTTATAACGGCGCAGGTGAGACCAACT	5700	6780	tatcaactTaaacacattttaccgtcaagTgattctctttccacatgaccaaactcgccatt	6839
Qy	5700	atttgacaaacaccccgcaagaacattgacagaataaaatgaacaaaggttatccgctct	5759	6781	TATCACTTAAACCAATTTTACCCTCAAGTGTCTCTTTTTCACCATGACCAATTCGCATT	6840
Db	5701	ATTTCACCAACAAACCCCGCAAGCCATTGACGAATAAATGAACAGGTTATCCGCTTCT	5760			



Qy 1560 ataaaccgagtcggttcaatgtgacacagctagaacggttggttaagtggctaaagagc 1619  
Dbb 1561 ATAAACCAGATCGCGTCAATGTGGCACAGCTAGAGCGGTGGTGAAGTGGCTAAGGAGC 1620  
Qy 1620 gtgaattactttcagtggtgatataacagctactgacgtataaaataggttttgataata 1679  
Dbb 1621 GTAGAAATTTACTTTTCAGGGTGATGATACAGTACTGAGCTAAAAATAGGTTTGGATAATA 1680  
Qy 1680 cttaactattaaaggttggtgcagagacaaacgcattaaacgataataataatcggttg 1739  
Dbb 1681 CTTTAACTATTAAAGGTGGTGCAGACACCAACGCATTAAACGATTAATAATATCGGTGTGG 1740  
Qy 1740 taaagagcgtgataatagtggtctgaagtttaaacttgctaaaccttaaacacttta 1799  
Dbb 1741 TAAAAGAGCGCTGATTAATAGTGGTCTGAAAGTTTAAACTTTGCTTAAACCTTTAAACAATCTTA 1800  
Qy 1800 ctgaggtgaatacaactacattaaatgccaacacacagtttaaggttagtgtagta 1859  
Dbb 1801 CTGAGGTGAATACAACACTACATTAAATGCGACAAACACACAGTTAAAGTAGGTAGTAGTA 1860  
Qy 1860 gtactacagctgaattattagtgatgtattaaacctttaccagcccaataacagcgatc 1919  
Dbb 1861 GFACTACAGCTGAATTAATGAGTGTAGTTTAACTTTACCCAGCCCAATACAGCGATC 1920  
Qy 1920 aaagcaagcaaaacccgtctatggtttaaagggtgaagttaacttaataatgcagaaa 1979  
Dbb 1921 AAAGCACAAAGCAAAACCGTCTATGGCGTTAATGGGGTGAAGTTTACTAATAATGAGAAA 1980  
Qy 1980 caacagcagcaatcgccactactcgtattaccagagataaaattggtttgctcgagatg 2039  
Dbb 1981 CAACAGCAGCAATCGGCACACTACTCTGATTATACAGAGATAAATTTGGCTTGGCTCGAGATG 2040  
Qy 2040 gtgatgttataaaacacagcaccatatttgataaaaaaacacttaagtgggtagtg 2099  
Dbb 2041 GTGATGTTGATGAAAAAACAAGCACCATATTTTGGATAAAAAACAACCTTAAAGTGGGTAGTG 2100  
Qy 2100 ttgcaattaccatagacaattgcatgtagtcagggtaataaaagatcagtaactcttgca 2159  
Dbb 2101 TTGCAATTTACCATAGACAATGGCATTTGATGCAGGTAAATAAAGATTCAGTAACTTGCCA 2160  
Qy 2160 aagtgacagctgaacgtcggttaccatcgacagctcaaacgcgcaagcctactt 2219  
Dbb 2161 AAGGTAGCAGTGTACGATGCGGTTACCATTGACACAGCTCAAGCGCCAGCGCTACTT 2220  
Qy 2220 taaacgagcgctgcatcagtggtcacacctactgaaatacagttgtagtgaagtg 2279  
Dbb 2221 TAAAGCAGCGCGTGGCATCAGTGTACACCTACTGAAATATACGTTGATGTGCTAAGAGTG 2280  
Qy 2280 gcaatgttacgcccccaacttaacatgtggtgaaacccagcagcttaacagtgatg 2339  
Dbb 2281 GCAATGTTACCGCCCCCAACTTAACATTTGGCGTGAACACCCAGCTTAACAGTGATG 2340  
Qy 2340 gcaatgtataaatttagtggttaaggttagtggtgcgaacaaatagcttagttaccgctg 2399  
Dbb 2341 GCACCTAGTATAATTTAGTGTAAAGGTAGTGTGACGAACATAGCTTAGTTACCGCGG 2400  
Qy 2400 acaatttggcaagctatctataatgaagtcgaatgcagcgctgacgtctctacaagct 2459  
Dbb 2401 AACATTTTGGCAAGCTATCTAAATGAAGTCAATGCAACCGCTGACAGTGTCTACAAAGCT 2460  
Qy 2460 ttaccgttaagaagaagacgatgatgaccccaacgctatcacgltggctlaaagatacga 2519  
Dbb 2461 TTACCGTTAAAGAAGAAGACGATGATGACCCCAACGCTATCACCGTGGCTTAAGATACGA 2520  
Qy 2520 caaaaaatgcggcgagtcagcatcttaaaactcaaaaggttaaaacggtctcaacggttg 2579  
Dbb 2521 CAAAAATGCGCGCGCAGTCAGCATCTTAAACTCAAAAGGTAAAAACGGTCTTAACGGTTG 2580  
Qy 2580 ctacccaaaaagatggttacggttacctttgggttaacgaagatacgggtctaccattg 2639  
Dbb 2581 CTACCAAAAAAGATGGTACGGTTTACCTTTGGCGTTTACCAAGATAGCGGTCTGACCAATTG 2640  
Qy 2640 gcaaaagcaccctaaacaacgatggcttgactgttaagatataccacgaacaaatccaag 2699

Dbb 2641 GCAAAAGCACCCTAAACACAGATGGCTTGACTGTTAAAGATACCAACGAACAATCCAAG 2700  
Qy 2700 tcsgtctaagtcgcatataattacttaattgaatgtagtaatccaggtactggcattg 2759  
Dbb 2701 TCGGTGCTAATGGCATTAATTTACTAATGTGAATGGTAGTANTCCAGGTACTGGCATTG 2760  
Qy 2760 caaatccgctgcattaccagagataaaattggctttgctgttctgtatggtgcagttg 2819  
Dbb 2761 CAATACCGCTCGCATTTACCAGAGATAAAATTTGGCTTTGCTGTGATGGTGCAGTTG 2820  
Qy 2820 atacaaacaaaccttctctgatcaagagaagctacaagttggcaatgttaagattacca 2879  
Dbb 2821 ATACAAACAACACCTTATCTTGTATCAAGACAAGCTTACAAAGTTGGCAATGTTAAAGTTACCA 2880  
Qy 2880 acactggcattaacgcaggttggttaagcccatcacaggtgtgtccccaacactgcctagca 2939  
Dbb 2881 ACATGCGATTACGCAAGTGGTAAAGCCATCACAGGCTGTCCCAACACTGCCTAGCA 2940  
Qy 2940 ttgcgcatcaaaagtagccgcaacatagaactgggcaatatacaatccaaagacaagaacaaat 2999  
Dbb 2941 TTGCGGATCAAAAGTAGCCGAACATAGAACTGGGCAATACAAATCCAAGACAAAGACAAAT 3000  
Qy 3000 ccaacgctgcagcattaatgatataataacagcgttttaacctataaaaaataaataca 3059  
Dbb 3001 CCAACGGCTGCCAGCATTAATGATATATTAATACAGGCTTTAACTTAAAAATAATAAACA 3060  
Qy 3060 acccattgactttgtctccacttatgacattgttgactttgccaatggcaatggcaca 3119  
Dbb 3061 ACCCATTTGACTTTGTCTCCACTTATGACATTGTGTGCTTTGCCAATGGCAATGCCACCA 3120  
Qy 3120 ccgcaacagtaaccatgatacccgtaacaaaacagtaaaagtgaatgttatatgatgtgaatg 3179  
Dbb 3121 CCGCACAGTAAACCCATGATACCGCTTAACAAAACAGTAAAGTGGTATATGATGTGTAATG 3180  
Qy 3180 tggatgatacaaacattcatctaacagcactgatgacaataaaaaacttggcgctcaaaa 3239  
Dbb 3181 TGGATGATACAAACCATTCATCTAACAGGCACTGATGACATAAAAAACTTGGCGTCAAAA 3240  
Qy 3240 ccacaaactgaacaaacaaagtgctaatggttaatacagcaactaaacttttaactgaact 3299  
Dbb 3241 CCACCAAACTGAACAAAACAAGTGCTAATGTTAATACAGCAACTAATCTTAAATGTTAACT 3300  
Qy 3300 ctagtgaagaatgccccttgttaacgcaaaagacatcgccgaataatctaaacacccctag 3359  
Dbb 3301 CTAGTGTATGAAGATGCCCTTTTAAAGCAAAAGACATCGCCGAAAAATCTAAACACCTAG 3360  
Qy 3360 ccaaggaattcacaccacccaaaggcagacagacacgcgcctacaacacctttaccgltta 3419  
Dbb 3361 CCAAGGAAATTCACACCACCAAGGCACAGCAGACACCGCCCTACAAAACCTTTACCGTTA 3420  
Qy 3420 aaaaggtagatgaataataatgctgatgacgcaacgcacacacccctgggtcaaaaga 3479  
Dbb 3421 AAAAGGTAGTGAATAATTAATATGCTGATGACCCCAACGCCATCACCGTGGGTCAANAAGA 3480  
Qy 3480 acqcaaaaatacaagtcacaccccttaacactcaaaaggtgaaaacggttcttaataataaaa 3539  
Dbb 3481 ACGCAATAATAACAGTCAACACCCCTAACACTCAAAAGGTGAAAAACGGTCTTAATATTAAAA 3540  
Qy 3540 ccgcaaaaaatgttcaggttacctttggcatttaacacacacagcggtctctaaagccgca 3599  
Dbb 3541 CCGACAAAATGGTACGGTTACCTTTGGCATTTAACACCAACAGCGCTTTAANGCGCGCA 3600  
Qy 3600 aaagcaccctaaacagcggtggttcttataaaaaacccactggtgagcgcaacaaatcc 3659  
Dbb 3601 AAAGCACCTTAACGACGCTGGCTTGTCTATTAAAAACCCCACTGGTAGCAACAAATCC 3660  
Qy 3660 aagtcggtgctgatggcgtagaagtttgcgaagtttaataataggtgttgtaggtgctg 3719  
Dbb 3661 AAGTCGGTGTGATGGCGTGAAGTTTGGCAAGGTTAATAATATGCTGTGTAGGTGCTG 3720  
Qy 3720 gcaatgaagcacaactcgcatlaccagagatgaatggctttactgggactaaaggct 3779



Db 3721 GCATTGATGGCACAACATCGCATACACAGATGAATGGCTTTACTGGGACTAATGGCT 3780  
 QY cacttgataaaagcaaacccaccctaaagcaaaagcagcgttaacagcaggtggtgataaaaga 3839  
 Db 3781 CACTTGATAAAGCAAAACCCACCCTAAGCAAAACACGCGATTAACGCGAGGTGGTAAAAAGA 3840  
 QY ttaccacattcaatcaggtgagatgccccaaagacagcatgatgctgtgacagggcgca 3899  
 Db 3841 TTACCACATTCATCAGGTGAATGGCCAAACAGCCATGATGCTGTGACAGCGGCA 3900  
 QY agatttatgattlaaaacacgaaactgaaacaaactcaagcagctactgccaaacacagcac 3959  
 Db 3901 AGATTTATGATTTAAACACGAACTTGAAACAAATAATCAGCAGTACTGCCAAACAGCAC 3960  
 QY aaactcattacagcaattctcagtagcagatgaacaaggaataaactttacggttagta 4019  
 Db 3961 AAAACTCATFACAGGAATCTCAGTAGCAGATGAACAAGGTAATAACTTTACGGTTAGTA 4020  
 QY acccttactccagttagacacctcaagacccctctgatctcactcttgagggtaaa 4079  
 Db 4021 ACCCTTACTCCAGTTATGACACCTCAAGACCTCTGATGTCATCAGCTTGCAGGTGAAA 4080  
 QY acggcattaccaccaaagtaataaagggtgtgtcggtggcattgaccaaaccacaaag 4139  
 Db 4081 ACCGCAATACCAACAGGTAATAAAGGTGTGTGGCTGTGGCCATTTGACCAACCAAG 4140  
 QY gcttaaccacgcctaagctgacgtgggtaataatggtgcaagcattgtcatlgaca 4199  
 Db 4141 GCTTAACCAACGCCTAAGTGACCGTGGGTAATAATAATGCAAGGCAATGTCTATTGACA 4200  
 QY gccaaatggtcaaaataccatcacaggactaaggcaacactctagctaaatggtaccaatg 4259  
 Db 4201 GCCAAATGTGTCAAAATFACCATCACAGGCTAAGCAACACTCTAGCTAATGTACCAATG 4260  
 QY ataaagtagcgtacgcaccacagacagggcaataataatcaagacagaaacacccc 4319  
 Db 4261 ATAAAGTAGCGTACGCCACACAGACAGGGCAATAATAAAGACAGACAAACCCC 4320  
 QY gtgccccagcattgtgtgctgaagcagcagctttaaacttgcaaggcaatgggtaag 4379  
 Db 4321 GTGCCCGCAGCATTTGTGATGTGCTAAGCGCAGGCTTTAACTTGAAGGCAATGGTGAAG 4380  
 QY cggttgacttgtctccacttagacacgtcaacttgcggtgcaatggcaccacgg 4439  
 Db 4381 CGGTTGACTTTGTCTCCACTTATGACACCGCTCAACTTGGCGATGGCAATGCCACCACCG 4440  
 QY ctaaggtgacctatgatgacacaaagcaaaaccagtaagtggtctatgatgtcaatgtgg 4499  
 Db 4441 CTAAAGTGACCTATGATGACACAAAGCAAAACCAAGTAAAGTGGTCTATGATGTCAATGTGG 4500  
 QY atgatacaaccattgaagtttaaagataaaaacttggcgtaaaaaccacacacattgacca 4559  
 Db 4501 ATGATACAACCATTTGAAGTTAAAGATAAAAAAATCTGCGGTAAAAACCAACCACTTGACCA 4560  
 QY gtactggcacaggtgtataataatttgcctaaagcaatacaagctactggcgatggcgttg 4619  
 Db 4561 GTACTGSCACAGGTGTATTAATAATTTGCCCTAAGCAATCAAGTACTTGGCGATGGCGGTTG 4620  
 QY tcaagggcagtgatagctgtgctcaactaaacaccttacttgcgacatccaaactgcca 4679  
 Db 4621 TCAAGGCCAGTGATATCGTTGCTCATCTAAACACCTTATCTGGCGACATCCAAACTGCCA 4680  
 QY aaggggcaagcgaagcaaacctcagcaggtctatgtggatgctgaaggcaataaagtgca 4739  
 Db 4681 AAGGGCAAGCCAAAGCAACACTCAGCAGGCTATGTGGGATGCTGATGGCAATAAAGTGCA 4740  
 QY tctatgacagtacgataacaaagtactatcaagccaaaataatgtggcaaggttgataaaa 4799  
 Db 4741 TCTATGACAGTACCGATAACAAGTACTATCAAGCCAAAAATGATGGCACAGGTTGATAAAA 4800  
 QY ccaaaaggttgcaagacaaactggtcgcccaagcccaaacccacagatggcacattgg 4859  
 Db 4801 CCAAAAGGTTGCCAAAGACAAACTGGTGCGCCCAAGCCCCAACCCAGATGGGCACATTTGG 4860

QY 4860 ctcaaatgaatgtcaaatcagttcatttaacaagaacaagttaaattgatgtccaaataaaagc 4919  
 Db 4861 CTCAAAATGAATGTCAAATCAGTCATTAAACAAGAACAAAGTAAATGATGCCAATAAAAAGC 4920  
 QY aaggcatcaatgaagacaacgccttctgttaagagacttgaaaaagcgcgtcttgataaca 4979  
 Db 4921 AAGGCATCAATGAAGACAAACGCCCTTTGTTAAAGACTTTGAAAAGCCGCTTCTGTATAA 4980  
 QY aaaccaaaacgcgcagtaactgtgggtgatttaaagcgttgcccaaacacccgctga 5039  
 Db 4981 AAACCAAAACGCCGCGAGTAACCTGGGTGATTTAAATGCCGTGGCCCAACACCGCTGA 5040  
 QY cctttgcaggggtatagcaggcacaacgcgttaaaaactggcgagactttgaccatcaaaag 5099  
 Db 5041 CCTTTGCAGGGGATACAGGCACAAACGGCTAAAAAAGCTGGCGAGACTTTGACCATCAAG 5100  
 QY gtgggcaaacagacacacaaataagcctaaccgataataacatcggtgtggtagcaggtactg 5159  
 Db 5101 GTGGGCAAAACAGACACCAATAAGCTAACCGATAATAACATCGGTGTGGTAGAGGTACTG 5160  
 QY atggcttcactgtcaaaacttgcaaaagacacctaaccacttaacagcgttaatgcaggtg 5219  
 Db 5161 ATGGCTTCACTGTCAAACTTTGCCAAAGACCTTAACCAATCTTAACAGCGTTTAATGCAGGTG 5220  
 QY gcaccaaaatgtgatgacaaaggcgtgtctttgttagactcaagcgtgcaagccaaagcaa 5279  
 Db 5221 GCACCAAAATGTATGACAAAGGGGTGTCTTTGTGTAGACTCAAGGGGTCAAGCCAAAGCAA 5280  
 QY aacccctgtctgaatgtgcaaatgggctggactgggtggcaagggtcatcagtaattggtg 5339  
 Db 5281 ACACCCCTGTGCTAAGTGGCAATGGGCTGGACCTGGGTGGCAAGGTCAATGTAATGTGG 5340  
 QY gcaaggcacaaagataccgacgctgccaatgtacacaggttaaacgaaagtagcgaact 5399  
 Db 5341 GCAAGGCACAAAGATACCGACGCTGCCAATGTACAAAGTTAAACAGAGTACGCAACT 5400  
 QY tgttggtctgtgttaatgtctgtgttaatgatacgtcgtgcaatcaggttaaacattgccc 5459  
 Db 5401 TGTGTGGTCTTGTGTAATGCTGTGTAATGATAACGCTGACGGCAATCAGGTAAACATTGCCG 5460  
 QY acatcaaaaagaccccaaatcagggttcatactaaacccgactgtcatcaaaagcagga 5519  
 Db 5461 ACATCAAAAAGACCCCAAAATTCAGGTTTCATCATCTAACCCGACCTGTTCATCAAGACGGCA 5520  
 QY cgttacttgcggttaaggttaataacgataccgaaaaacttgccactgggtggtatacaag 5579  
 Db 5521 CGGTACTTGGCGGTAAAGGTAAATTAACCGATACCGAAAAAATTCGCACCTGGGTGATACAG 5580  
 QY tgggctgtgataaagacggtcaacgcgtacgcgtatttaagcaatgtttggttcaaaaccc 5639  
 Db 5581 TGGCGGTGGATTAAGACGGCAACGCTAACGGCGATTTAAGCAATGTTTGGGTCAAAAACCC 5640  
 QY aaaaagatggcagcaaaaaagccctgctgccactataacgcgcaggtcagaccact 5699  
 Db 5641 AAAAAGATGGCAGCAAAAAAGCCCTGCTCGCCACTTATAACGGCGCAGGTGAGACCAACT 5700  
 QY attgacaaacaccccgagagccattgacagaaataatgaacaggtatccgctctct 5759  
 Db 5701 ATTTGACCAACACCCCGCAGAACCAATTACAGAAATAATGAACAGGTATCCGCTTCT 5760  
 QY tccatgtcaacgatggcaatcaagagcctgtgtgtaagggcgtaacggcattgactcaa 5819  
 Db 5761 TCCATGTCAACAGATGGCAATCAAGAGCCTGTGGTACAAGGGCGTAACGGCATTGACTCAA 5820  
 QY gtgcctcaggcaagcactcagtggtgtagtttccaggccaaagggcagatggtgaagccg 5879  
 Db 5821 GTGCCTCAGGCAAGCACTCAGTGGCGATAGTGTTCAGGGCCAAAGCGAGATGGTGAAGCCG 5880  
 QY ccgttgccatagcgagacaaaccccaagcaggaacccatccatcgccatcggtgataacg 5939  
 Db 5881 CCGTTGCCATAGGCAGACAAACCCCAAGCAGCAACCAATCCATCGCCATCGGTGATAACG 5940





QY	1368	tacaggcagatgtgccttaggctcagggttcgtctgctcactcagagtgataataattcta	1427	Db	1742	CTTACAAAGCTTTACCGTTAAAGAAAGACGATGATGACGCCAACGCTATCACCTGG	1801
Db	662	TACAGGCGCAGTATGGCCTTAGGTCAGGTTCTGTGTGCTACTCAGAGTGATAAATCTTA	721	QY	2508	ctaaagatcagcaaaaaatgcgcgcgcagctcagcagctcttaaaactcaaggtaaaacg	2567
QY	1428	gacgcgcctatacacaataaccaggcactagaccacccagttccaagccaccacaataa	1487	Db	1802	CTAAAGATAGACANAAATGCGCGCAGTCAGCATCTTAAAACTCAAGGTAAANACG	1861
Db	722	GACCGGCTATATACANNAATACCCAGGCAGTACGCCCAAGTTTCAGGCCACCAANTA	781	QY	2568	gtctaacggttgctaccacaaaaagatgtagcggttacctttggccttagcacaagatagc	2627
QY	1488	cgaagcggtgctcactttccatttgtagtaacctctatcaaacgtaaaaatcatcaatgctg	1547	Db	1862	GTCTAACGGTTGCTTACCAAAAAGATGTCAGGTTACCGTTTGGCTTAGCAAGTAGCG	1921
Db	782	CGAAGCGGGTCCACTTTTCCATTTGGTAGTAACCTCTATCAAAAGTAAATCATCAATGTCG	841	QY	2628	gtctgaccattggcaaaagcacccttaaaacacgatggtcttgactgtttaagataaccaacg	2687
QY	1548	gtcagagtgtaataaaaacagatgcgggtcaatgtgcacagctagaaagcggttgtaagt	1607	Db	1922	GTCTAACGGTTGCTTACCAAAAAGATGTCAGGTTTACCGTTTGGCTTAGCAAGTAGCG	1981
Db	842	GTCAAGGTGTTAATAAACCGATGCGGTCAATGTGTCACAGCTAGAACGCGGTGTGAAGT	901	QY	2688	acaaatccaaagtcggtgctaaatggtcattaaattacttaattgtaattgtagttagtaacag	2747
QY	1608	gggctaaggagcgtagaattacttttcagggtgatataacagtagctagcgtaaaaatag	1667	Db	1982	ANCAATCCAAAGTCGGTGAATGGCATTAATTTACTAATGTGAATGGTAGTAATCCAG	2041
Db	902	GGCTAAGGAGCGTAGAATTTACTTTTCAGGGTGATGATAACACTACTGACGTAANATAG	961	QY	2748	gtactggcattgcaaataccgctcgcatcaccagagataaaattggcttggcttggcttctg	2807
QY	1668	gtttggataatactttaactattaaagggtgtgcagagaccaacgcattaccggataata	1727	Db	2042	GTACTGGCATTTGCAATACCGCTCGCATTACCAGAGATAAAATTTGGCTTTGCTGTTCTG	2101
Db	962	GTTTGGATAATACTTTAACTATTAAAGGTGTCAGAGACCAACGGCATTAACCGATAATA	1021	QY	2808	atggtgcagttgatatacaaaacacaccttatcttgatcagaagcaagctacaagttggcaatg	2867
QY	1728	atcgggtggtgtaaaagagcgtgataataatgtggtcgtgaagttaaaacttgctaaaaact	1787	Db	2102	ATGGTGCAGTTGTATACAAACAAACCTTTATCTTGATCAAGACAAAGCTCAAGTTGGCAATG	2161
Db	1022	ATATCGGTGTGTTAAAGAGCGCTGATAATAGTGTCTCTGAAGTTAAACTTTGCTAAAACTT	1081	QY	2868	ttaagattacaaacactggcattaaacgcaggtggtgaagccatcacaggcgtgcctccaa	2927
QY	1788	taacaactctactgaggtgaaatacaactacattaaatgcccacaacacacagttaaagtag	1847	Db	2162	TTAAGATTACCAACACTTGGCATTAACGCAGGTGTTAAGCCNTCACAGGGCTGTCCCAA	2221
Db	1082	TAAACAATCTTACTGAGTGAATACAACTACATTTAATGCCACCAACACAGTTAAGTAG	1141	QY	2928	cactgcctagcattgcccagatcaaaagtagccgcaacatagaaactgggcaatacaatccaa	2987
QY	1848	gtagtagtagtagtactacagctgaattatttgagtgatagtttaaacctttaccagccca	1907	Db	2222	CACTGCCATGTCATTTGCCGATCAAAAGTAGCCCAACATAGAACTGGCAATACAATCCAAG	2281
Db	1142	GTAGTAGTAGTAGTACTACAGCTGAATATTAGTGTGATAGTTTAACTTTACCAGCCCA	1201	QY	2988	acaaagcaaatccaaagcgtgcagcgaattaatgatatataaaatacaggcgtttaaccta	3047
QY	1908	atacaggcagtaaaagcaagaacacccgtctatggcgttaattgggtgaagttaacta	1967	Db	2282	ACAAAGCAAACTCAACGCTGCCAGCATTAATGATATATTAAATACAGGCTTTAACCTAA	2341
Db	1202	ATACAGCAGCTCAAGCACAAAGCAAAACCCGTCTATGCGGTTAATGGGGTGAAGTTTACTA	1261	QY	3048	aaaataaacaacccccattgactttgtctccacttatgacattgttgacttttgccaatg	3107
QY	1968	ataatgcagaacaacagcagcaatcgccactcgtattaccagagataaaattggct	2027	Db	2342	AAATATATACACCCCATTTGACTTTGTCTCCACTTATGACATTTGTTGACTTTGCCAATG	2401
Db	1262	ATAATGCAGNAACACAGCAGCAATCGGCACACTCTCGTATTACAGAGATAAATTTGGCT	1321	QY	3108	gcaatgcacacccgcgcacagtaaccctatgataccgctaaacaaaccagtaaaagtgtat	3167
QY	2028	ttgctcgagatggtgtgtgttgataaaaaacaagcacccatatttgggataaaaaacaacta	2087	Db	2402	GCAATGCCACACCGCCACAGTAACCCCATGATACCGCTAACAAAACCAAGTAAAGTGTAT	2461
Db	1322	TTGCTCGAGATGCTGATGTGTGATGAAAAACAAGCACCATATTTTGGATAAAAAACAACCTTA	1381	QY	3168	atgatgtgaatgtgtagatatacaaacattcatctaaacaggcactgatgacaataaaaaac	3227
QY	2088	aagtggtagtgtgtaattaccatagacaatggcattgtagtcaggttaataaaaaagatca	2147	Db	2462	ATGATGTGAATGTGGATGATACAAACCATTCATCTAACAGGCACCTGATGACAAATAAANAAC	2521
Db	1382	AAGTGGTAGTGTGCAATTTACCATAGACAAATGGCATTTGATGAGGTAATTAANAAGATCA	1441	QY	3228	ttggcgtcaaaacaccacaaactgaacaaagtgctaaatggtaatacagcaactaaact	3287
QY	2148	gtaactctcgaaggttagcagtgctaacgatcggtgtaccatcgcaacagctcaaaagcg	2207	Db	2522	TTGGCGTCAAAACCCACCAACTGAACAAACAAAGTGTCTAATGGTAATACAGCACTACT	2581
Db	1442	GTAACTTTGCCAAAGGTAGCAGTGTCAACGATGCGGTTACCTTCAGACAGCTCAAGCCG	1501	QY	3288	ttaatgttaactctagtagtaagatgccccttgtaacgcgaagacagatcgccgaatac	3347
QY	2208	ccaagcctaactttaaacgcgagcgctggcatcagtgctcacacctactgaaatatcagttg	2267	Db	2582	TTAATGTTAACTCTAGTGTAGTGAAGATGCCCTTGTAAACGCCAAAAGACATCGCCGAAATC	2641
Db	1502	CCAAGCCTACTTTAAACGCCAGGCGCTGGCATCAGTGTACACCTTACTTGAAATATCAGTTG	1561	QY	3348	taaaaccccttagcgaaggaataatcaaccccaaaaggcacagacagacacccctataaaa	3407
QY	2268	atgctaagagtggaatgtttacgcgccccaaacttaacaattggcggtgaacacacacagc	2327	Db	2642	TAAACACCCCTAGCCAAAGGAATTCACACCAACCAAGGCACAGCACCCCTTACANA	2701
Db	1562	ATGCTAAGAGTGGCAATTTTACCGCCCCAACTTACAACATTGGCGTGAACACCCAGC	1621	QY	3408	ccttaccgtttaaaggttagatgaaaaataatgctgtagcgccaacgcccatcacg	3467
QY	2328	ttaacagtgtggaactagtgataaaatttagtgttaaggttaggttacgaacaatagct	2387	Db	2702	CCTTTACCGTTAAANAGGTAGATGAANAATAATATGCTGATGACGCCCAACGCCATCACCG	2761
Db	1622	TTAACAGTGTGCGACTAGTGATTAANAATTTAGTGTTAAGGGTAGTGTGACGAACATAGCT	1681	QY	3468	tgggtcaaaagacgcgaataataatcaagtcacaacccctaactcaaaaggtgaacacggtc	3527
QY	2388	tagttaccgcgcgaacatttggcgaagctatctaaatgaagtcaatcgcaacggtgacagtg	2447	Db	2762	TGGGTCAAAAGACGCAAAATAATCAAGTCAACACCCCTAACACTCAAAAGGTGAAACCGTC	2821
Db	1682	TAGTTACCGCCGAACAATTTGGCAAGCTATCTAAATGAAGTCAATCGAACGCTGACAGTG	1741	QY	3528	ttaatatataaacacgcgaacaaaatgtagcggttacctttggcattaaacacacacagcggtc	3587
QY	2448	ctctacaaagcctttaccggttaaaagaagaacgatgatgacgcccaacgctatcacgcgtg	2507				





QY 1022 ttgcttgatcagatggttaattccttaacatcccgaaaggtactctctgattaacgatcttatt 1081  
D 481 TTGCTTGATCAGCATGGTAATCTTAAACATCCGAAAGGTACTCTGTGATTAACGATCTTATT 540  
QY 1082 aacggccatcagtatataaagaataacgaagctcaagataatgatgtataaatataga 1141  
D 541 AAGGCCATGCGATGATTTAAAGAAATACGAGCTCAAGGATATGATGTAAATATAGA 600  
QY 1142 cgcacaaacgcaagcgagacacgcagctactgacgtgagagcattgtcatatgcacaggg 1201  
D 601 CGCACAAACGCAAGCGGACACCGCCAGTACTGACGTGGGAGCCATGTATATGACAGGGT 660  
QY 1202 catttttccacgcctttggttacacgggcaacagactaaagtgccattctctggcagtg 1261  
D 661 CATTTTCCAAACGGCTTTGGTACACGGGCAACAGCTAAAGTGCCCTATTCTCTGGCAGTG 720  
QY 1262 ggtcttgccgcacagcgcaggggccaactctacaatcgctattggttctctgatgcaacatct 1321  
D 721 GGTCTTGGCCGACAGCGGAGGCCAATCTACAATCGCTATTGGTCTGTGATGCAACATCT 780  
QY 1322 agctggtggagcgatagcccttggtgcaggtactcgtcagctacagggcagttatt 1381  
D 781 AGCTCGTTGGGAGCGATAGCCCTTGGTCCAGGTACTCGTGTCTAGCTACAGGGCAGTATT 840  
QY 1382 gccctagatcaaggtctctgtcactcagagtgataataattctagaccggcctataca 1441  
D 841 GCCCTTAGGTCAAGGTTCTGTTGCTACTCAGAGTGATAATTTCTAGACCGGCCCTATACA 900  
QY 1442 ccaataccccggcactagaccccccaagtttcaagccacccaataatacaagaagcggtcca 1501  
D 901 CCNAATACCCAGGCACTAGACCCCAAGTTTCAAGCCACCAATAATACGAAGCGGGTCCA 960  
QY 1502 ctttccattggttagtaactctatacaacgttaaaatcatactgctcggtgcaggtgtta 1561  
D 961 CTTTCCATTTGGTAGTAACATCTATCAAACTGTAATAATCAATCAATGCGTGACAGGTGTTAA 1020  
QY 1562 aaaaacgatcggttcaattgtgcacagctagagcggtggtgaaagtgggcaagagagcgt 1621  
D 1021 AAAACCGATGCGTCAATGTGGCACAGCTAGAAAGCGGTGGTCAAGTGGGCTAAGGAGCGT 1080  
QY 1622 agaattactttcaggtgatataacagctactgagctaaataagtttggatataact 1681  
D 1081 AGAATTTACTTTTCAGGGTGATGATAACAGTACTGACGTAAATAATAGTTTGGTAAATACT 1140  
QY 1682 ttaactattaaaggtggtgcagagacccaacgcatataaccgataataatcgggtggta 1741  
D 1141 TTAACATTTAAAGGTGGTCAGAGACCAACGATTAACCGATTAATAATATCGTGTGTA 1200  
QY 1742 aaagagctgataatagtggtctgaaagttaaacttgctaaacitttaacaacttact 1801  
D 1201 AAAAGAGCTGATAAATAGTGGTCTGAAAGTTAAACTTGGCTTAAACACTTTAAACAACTTACT 1260  
QY 1802 gaggtaatacaactacattaaatgccacaacacagttaaaggtaggttagtagtagt 1861  
D 1261 GAGGTGAATACAACTACATTAATGCGCACCAACACACAGTTAAGGTAGGTAGTAGTAGT 1320  
QY 1862 actacagctgaattatttgatgtagtttaacoccttaccagocccaatacacagcgatcaa 1921  
D 1321 ACTACAGCTGAATTAATTAGTGTAGTTTAAACCTTTTACCAGGCCAATAACAGGCAGTCAA 1380  
QY 1922 agcacaagaacacccgtctatgctgttaatgggtgaaagtttacttaataatgcagaaca 1981  
D 1381 AGCACAAGCAAAACCGTCATGCGGTTAATGGGGTGAAGTTTACTAATAATGCAGAAACA 1440  
QY 1982 acagcagcaatcgccactactcgtattaccagagataaaattggccttgcctogagatggt 2041  
D 1441 ACAGCAGCAATCGCACTACTCTGTAATTACAGAGATAAAATTTGGCTTTGCTCGAGATGGT 1500  
QY 2042 gatgttgataaaaaaacgacacatatttgataaaaaaacacttaagtggttagtt 2101  
D 1501 GATGTTGATAAAAAACAGCACCATATTTCGATAAAAAAACCTTAAGGTGGGTAGTGT 1560  
QY 2102 gcaattaccatagacaatgscattgagcaggtaataaaaaagatcagtaattcttgcaca 2161

D 1561 GCAATTTACCATACACAATGGCATTGTATGAGGTAAATAAAAGATCAGTAATCTTGCACAA 1620  
QY 2162 ggtagcagtgctaaacagatgcggtttaccatcgaaacagctcaaaagcccaagcctacttta 2221  
D 1621 GGTAGCAGTGCTTAACGATCGGTTACCATCGAACAGCTCAAAAGCCCAAGCCCTACTTTTA 1680  
QY 2222 aacgcaggcgctggcatcagtgccacacctactgaaatatcagttgctgaagagtgcc 2281  
D 1681 AACGCAGCGCTGGCATCAGTGTCACACCTTACTGAAATATACGTTGATGATGATGATGATG 1740  
QY 2282 aatgttaccgcccacttacaacattggtgtgaaacacacccagcttaacagtgatggc 2341  
D 1741 AATGTTACC GCCCAACTTACAACATTTGGGTGAAACCCACCGAGCTTAACAGTGATGGC 1800  
QY 2342 actagtgataaatttagtgtaaggttagtggtacgaacaatagcttagttacccgcgaa 2401  
D 1801 ACTAGTGATTAATTTAGTGTAAAGGTAGTGTGACGAAACATAGCTTAGTTACCGCGGAA 1860  
QY 2402 catttggcaagctatctaaatgaagtcacacggtgacagtgctctacaaaagcttt 2461  
D 1861 CATTTGGCAAGCTATCTTAAATGAAGTCAATCGAACGGCTGACAGTGTCTTACAAAGCTTT 1920  
QY 2462 accgttaaaagaagaacgatgatgacgccaacgctatcacctggtgctaaagatacagaca 2521  
D 1921 ACCGTTAAAGAAAGAGAGATGATGACGCCACACGCTATCACCGTGGCTAAGATACGACA 1980  
QY 2522 aaaaatgccggcgagtcagcatctttaaactcaaaaggtaaaaacgctctaaacggttgc 2581  
D 1981 AAAAATGCCGGCGAGTCAGCATCTTAAACTCAAAAGGTAAACACGGCTTAACCGGTTGCT 2040  
QY 2582 accaaaaagaatggtacgggttacctttggtcttagccaagatagcgtctcgaaccattggc 2641  
D 2041 ACCAAAAAGATGTTAGCGTTTACCTTTGGGCTTAGCCAAAGATAGCGGTCTGACCACTTTGGC 2100  
QY 2642 aaagacccttaaacacagatggcttgactgttaaagataccacagcaacaaatccaaagtc 2701  
D 2101 AAAAGACCCCTTAACACAGTGGCTTGACTGTAAAGATACCAACGAAACAAATCCAAGTC 2160  
QY 2702 ggtgctaattggtcatttaatttactaaatgtaattggtgtaattccaggttactggtgata 2761  
D 2161 GGTGCTAATGGCATTAATTTACTTAATGTGAATGGTGTAGTAATCCAGGTACTGGGATTTGCA 2220  
QY 2762 aatccgctcgatcattaccagagataaaattggttctggtctctgagtggtgagtgat 2821  
D 2221 AATACCGCTCGCATTTACCGAGATATAAATTTGGCTTTGCTGTTCTGTAGTGGTGCAT 2280  
QY 2822 acaaaaaaccccttatcttctgatcaagacaaagctacaagttggcaatgttaagattacaaac 2881  
D 2281 ACAAAACAACTTATCTTTGATCAAGACAAGCTACAAGTTGGCAATGTTAAGATTACCAAC 2340  
QY 2882 actggcataaacgcaggtggtggttaagccatcacaggtctccccaacactgcctagcatt 2941  
D 2341 ACTGGCATTAACGCGAGGTGGTAAAGCCATCACAGGGCTGTCCCCNACACTGCCTAGCAT 2400  
QY 2942 gccgatcaagtagcgcgaacataagaactgggcaatacaatccaaagacaaagacaaatcc 3001  
D 2401 GCCGATCAAAAGTAGCGGCAACATAGCACTGGGCAATACAATCCAAAGACAAACAAATCC 2460  
QY 3002 aacgctgcagcatttaattgatatataataacaggtcttaacctaaaaataataaacac 3061  
D 2461 AACGCTGCCAGCATTAAATGATATATTAATACAGGCTTTAACCTTAAATAATTAACAAAC 2520  
QY 3062 cccattgacttggctccacttatgacattgttgactttggcaatggcaatggccaccacc 3121  
D 2521 CCCATTGACTTTGTCTCCACTTATGACATTTGTGACTTTGCCAATGGCAATGCCACCACC 2580  
QY 3122 gccacagtaaacctatgatacccttaacaaaaacagtaaaagtgggtatgatgtgagtgtg 3181  
D 2581 GCCACAGTAACCCATGATACCGCTAACAAAACAGTAAGAGTGGTATATGATGTGAATG 2640  
QY 3182 gatgatacaacacattcatcacaagcactgatgacaataaaaaaacttggtcgtcaaaacc 3241

Db 2641 GATGATACAAACCAATTCATCTACAGGCACTGATGACAAATAAAAAAATTGGCGTCAAAACC 2700  
QY 3242 accaaactgaacaaaacagtgctaataatggttaatacacagcaactaacctttaatgttaactct 3301  
Db 2701 ACCAACTGAACAAAACAAGTGCTAATGGTAAATACAGCAACTAACTTTAATGTTAACTCT 2760  
QY 3302 agtgatgaagatgccttgtttaagcgaagaagacatcgcgaaataataaacacccctagcc 3361  
Db 2761 AGTGATGAAGATGCCCTGTGTTACGGCAAGACATCGCGAAATCTAAACACCCTAGCC 2820  
QY 3362 aaggaattcacaccaccacaaagccagcagacacacgcgcctcacaaacctttaccgttaa 3421  
Db 2821 AAGGAATTCACACCAACCAAGGCACAGACACCGCCCTACAAACCTTTACCGTTAA 2880  
QY 3422 aaggtatgaataataatgctgatgacgccaagccatcacccgtgggtcgaagaac 3481  
Db 2881 AAGGTAGATCAAAATAATAATGCTGATGAGCCCAACGCCATCACCGTGGTCAAAAGAAC 2940  
QY 3482 gcaataataaagtcacaccccttaacactcaaaaggtgaaacacggttottaatatataaac 3541  
Db 2941 GCAAAATAATCAAGTCAACACCCCTAACACTCAAAAGGTGAAACCGGTCTTAATATTA AAC 3000  
QY 3542 gacaaaaatggtacggttacctttgcatcattacacacacagcggtcttaaacgcgcga 3601  
Db 3001 GACAAAAATGTTACGTTACCTTTGGCATTTAACCAACAGCGGTCTTTAAAGCCGGCAA 3060  
QY 3602 agcacctaaacagcgttgctgtctattataaaaccccactgtagcgaacaaatccaa 3661  
Db 3061 AGCACCTAAACGACGCTGCTGTTCTATTAAAAACCCCACTGCTAGCGAACAATCCAA 3120  
QY 3662 gtcggtgctgatggcgtgaagtttgccagggttaataataatggtgttgatgctggc 3721  
Db 3121 GTCGGTGTGATGGCGTGAAGTTTGCCAGGTTAATAATGCTGTTGTAGTGTGCTGCC 3180  
QY 3722 attgatggcacaactcgattaccagatgaattggctttactggactaatgctca 3781  
Db 3181 ATTGATGGCACTCGCATACAGATGAATTTGGCTTTTACTGGGACTAATGGCTCA 3240  
QY 3782 ctgtataaagcaaacccacccctaaagcaagcggcattaacgcaggtgggtgaagaagatt 3841  
Db 3241 CTTGATAAAAGCAAAACCCCACTTAAGCAAGACGGCATTAACGCGTGTGTTAAAGATT 3300  
QY 3842 accaaattcaatcaggtgagatggcccaaaacagccatgatgctgtgacagcggcgaag 3301  
Db 3301 ACCAACTTCAATCAGGTGAGATTGCCCAAAACAGGCATGATGCTGTGACAGCGGCAAG 3360  
QY 3902 attatgatttaaaacccgaactgaaacaaataacagcagctactgccaaaacagcaca 3361  
Db 3361 ATTATGATTTAAACCCGAACCTGAAACAAATAACAGCACTGCTGCTGCAAAACAGCACAA 3420  
QY 3962 aactcattacagaattctcagtagcagatgaacaaggtaataaactttacggttagtaac 4021  
Db 3421 AACTATTACACCAATTCACGATGAGATGAACAAGGTAATAACTTTACGGTTAGTAAC 3480  
QY 4022 ccttactcoagttatgacacctcaagaccctctgatgtcatcacccttgcaggtgaaac 4081  
Db 3481 CCTTACTCCAGTTATGACACCTCAAGACCTCTGATGTCATCACCTTTGCAGGTGAAAC 3540  
QY 4082 ggcattaccaccaaagtaataaggtgtggtgctgtggtgctattacacaaacaaagc 4141  
Db 3541 GGCATTACCAACCAAGGTAAATAAAGGTGTGGTGGGTGGGCTATTGACCAACCAAGGC 3600  
QY 4142 ttaaccacgcctaagctgaacgtgggttaaataatggcaaaagccattgtcattgacagc 4201  
Db 3601 TTAACCAACGCCTAAGCTGACCGTGGGTAAATAATGGCAAGGCATTGTTCATTGACAG 3660  
QY 4202 caaaatggtcaaaataaccatcacagactaagcaacactctagctaatgttacaatgat 4261  
Db 3661 CAAAATGGTCAAAATACCATCACAGGACTAAGCAACACTCTAGCTAATGTTACCAATGAT 3720  
QY 4262 aaaggtagcgtacgcaccacagacagggcaataataataacaaagcaagacaaacccgt 4321  
Db 3721 AAAGGTAGCGTACGCACCAAGAACAGGGCAATATAATCAAGAGCAAGAACAAACCCGT 3780

QY 4322 gcgcgcagcattgtgatgtgtaagcgcaggttttaacttgcgaagcaatggtgaagcg 4381  
Db 3781 GCCCCAGCACTTGTGATGTGCTAAGCGCAGGCTTTAACTTGAAGCAATGGTGAAGCG 3840  
QY 4382 gttgaacttgcctccacttatgacacccgtcaactttgcgatggcaatgcaccacccgt 4441  
Db 3841 GTTGACTTTGCTCTCCACTTATGACACCGCTCACTTTGCCGATGCAATGGCACCCGCT 3900  
QY 4442 aaggtaccttatgatgacacaagaacacacagtaaaagtgtcttatgtatgtaattgtgat 4501  
Db 3901 AAGGTGACCTTATGATGACACAGCAAAACCACTTAAGTGGTCTATGATGTAATGTGGAT 3960  
QY 4502 gatacaaccttaagttaaagataaaaaacttggcgtaaaaaacacacacattgaccagt 4561  
Db 3961 GATACAACCACTTGAAGTTAAAGATAAAAACTTGGCGTAAAAACCCACCATTTGACCAGT 4020  
QY 4562 actggcaacaggttctaataaatttgcctaaagcaactcaactactggcgatgagcttctc 4621  
Db 4021 ACTGGCACAGGTGCTAATAAATTTGCCCTAAGCAATCAAGCTACTTGGCGATGGCGTTGTC 4080  
QY 4622 aagggcagtgatcgttgcctcatcctaacaacaccttatctgcgacatccaaactgcga 4681  
Db 4081 AAGGCCAGTGTATGCTGTGCTCATCTTAACACCTTATCTGGCGACATCCAAACTGCCAAA 4140  
QY 4682 ggggcaagccaaagcaaaactcagcaggtctatgtggtatgtgatggtgcaataaagtcctc 4741  
Db 4141 GGGCAAGCCCAAGCAACAACTCAGCAGGCTATGTGCTGCTGATGCAATAAGTGCATC 4200  
QY 4742 tatgacagtcacagatacaaaagtactatcaagcgaataatgatggcacagttgataaaac 4801  
Db 4201 TATGACAGTACCGATACAAAGTACTATCAAGCCAAAAATGATGGCACAGTTGATAAAC 4260  
QY 4802 aaagaagtgcgaagcaaaactggtcgccaaagcccaaacccagatgcacattgct 4861  
Db 4261 AAAGAAGTTCGAAGACAACTGCTGCCAAGCCCAACCCCGATGGCACATTGGCT 4320  
QY 4862 caaatgaattcgaatcagtcattacaagaacagtaaatgatgccaatacaaaaagca 4921  
Db 4321 CAAATGAATGCAATCAGTCAATTACAAAGAACAACTAAATGATGCAATAAAGCA 4380  
QY 4922 ggcataataagaacacgccttgttaaggactgaaaaagccgcttctctgatacaaa 4981  
Db 4381 GGCATCAATGAAGACACGCGCTTTGTTAAAGGACTTGAAAAAGCCGCTTCTGTATAACAA 4440  
QY 4982 accaaaaacccgcagtaactggtggtgatttaaatccggttgcccaaacacgcgtgacc 5041  
Db 4441 ACCAAAAACCCGCGTAACTGTGGGTGATTTAATGCCGTGGCCCAACACCCGCTGACC 4500  
QY 5042 ttltgcaggggatcacaggcaacacgcgctaaaaaactggcgagactttgaccatcaaaaggt 5101  
Db 4501 TTTGCAAGGGGATACAGGCACACGCGTAAAAAACTGGCGGAGACTTTGACCATCAAAAGGT 4560  
QY 5102 gggcaaacagacaccaataagctaaaccgataataaacctcggtgtggtgtagcaggtactgat 5161  
Db 4561 GGGCAACACAGACACCAATAAGCTAAACCGATAATAACATCGGTGTGTGTGAGCAGGTACTGAT 4620  
QY 5162 ggcctcactgtcaaaactgcgaagacacacaaacttaacacgcttaataatgaggtggc 5221  
Db 4621 GGCTTCTACTGTAAACTTGCCAAAGACTAACCAANTCTTAACAGCGTTAATGCAAGGTGGC 4680  
QY 5222 accaaaaattgatgacaagcggtgtctttttagactcaagcggtcaagcgaacaaagcaaac 5281  
Db 4681 ACCAAAAATTGATGACAAGGCGGTGCTTTTGTAGACTCAAGCGGTCAAGCCCAAGCAAC 4740  
QY 5282 acccctgtgctaagtgcaatggcgctggacccctggcgcaaggttcacgttaattgtgggc 5341  
Db 4741 ACCCCTGTGTAAGTGCATTAAGTGGCTGGACCTGGCGTGGCAAGGTCTATCAGTAATGTGGGC 4800  
QY 5342 aaaggcacaagaatcacgcagcgtccaatgtacacagittaaacgaagtcacgaacttg 5401  
Db 4801 AAAGGCACAAAGATACCGACGCTGCCAATGTACACAGTTAAACAGGAAGTACGCAACTTG 4860

Qy	5402	ttgggtcttgtaatactggttaataagcctgaacggcctacaggcaatcaggtaaacattgccac	5461		
Db	4861	TTGGGCTCTGTGTAATGCTGTGTAATCATACAGCTACCGCAATCAGGTAAACATTTGCCGAC	4920		
Qy	5462	atcaaaaaagcccaaaatccaggttcatcatcaaccgcactgtcatcaaaagcaggcacg	5521		
Db	4921	ATCAAAAAGACCACAAATTCAGGTTTCATCATCTAACCGCAGCTGTCATCAAGACGGCAGC	4980		
Qy	5522	gtacttggcgttaaggttaataacgataccagaataaattgccactggttgatatcaagtg	5581		
Db	4981	GTACTTTGGCGGTAAGGTAATTAAGCATACCGAAAAACTTGGCACTGGTGTATACAAAGTG	5040		
Qy	5582	ggcgtgataaagacgcacacgcctaaacggcgatttaagcaatgtttgggtcaaaacccaa	5641		
Db	5041	GGCGTGGATAAAGACGGCAACGCCTAACGGCCATTTAACCAATGTTGGGTCAAAACCCAA	5100		
Qy	5642	aaagatggcagcaaaaaagccctgtcgcacctataacgcgcgaggttcagaccacatat	5701		
Db	5101	AAAGATGGCAGCAAAAAGCCCTGCTCGCCACTTATAACGCGCAGGTCAGACCAACTAT	5160		
Qy	5702	ttgacaacaaccccgagaagccattgacagaataaataagcaaggtatccgctcttc	5761		
Db	5161	TTGACCAACAACCCCGCAAGCCATTGACAGATAAATGAACAAAGGTATCCGCTTCTTC	5220		
Qy	5762	catgtcaacgatggcaatcaagagcctgtgtacaaaggcgttaacggcattgactcaagt	5821		
Db	5221	CATGTCACGATGGCAATCAAGAGCCTGTGTGTACAAAGGCGTAACGGCATTGACTCAAGT	5280		
Qy	5822	gcctcaggcaacatcagtggtggttagtttccaggccaaggcagatggtgaagcgc	5881		
Db	5281	GCCTCAGGCAAGCACTAGTGGCGATAGTGTTCAGGCAAGGCAGATGGTGAAGCCGCC	5340		
Qy	5882	gttcctataggcagacaacccaacgacgcaaccaatccatcgccatcggtgataacgca	5941		
Db	5341	GTTCCTCATAGGCAGACAACCCACGAGCAACCAATCCATCGCCATCGGTGATACGCA	5400		
Qy	5942	caagccacgggcgataatccatcgccatcggtgacaggaatgtggtgagcaggttaagcac	6001		
Db	5401	CAAGCACGGCGATCAATCCATCGCCATCGGTACAGGCAATGTGTAGCAGGTAAAGCAC	5460		
Qy	6002	tctgtgcatcagcgaccccaagcactgttaaggtgataacagttacagttggtgta	6061		
Db	5461	TCGTGTGCTATCGGCGACCCNAGCACTGTATAGGCTGATTAACAGTTACAGTTGGTAT	5520		
Qy	6062	aacaacagttaccgatgccactcaacacccgatgttcttgggtggtggtgcaataaacatcac	6121		
Db	5521	AACAACCAAGTTTACCAGTCCACTCAACCCGATGTCTTTGGTGTGGGCAATAACATCAC	5580		
Qy	6122	gtacccgaaagttaactcgtgtgctttagttcaaaccttgccatcagtgacggcacacac	6181		
Db	5581	GTACCCGAAAGTAACTCGGGTGCTTTAGGTTCAAACTCTGCCATCATAGTCAGGCACACAC	5640		
Qy	6182	gcaggcacacaagcccaaaaaatctgcgcgcagcagcaggttacaccaccacagcagtgca	6241		
Db	5641	GCAGGCACACAGCCAAAATCTGACGCACAGCAGGTACAAACCACCAAGCAGGTGCA	5700		
Qy	6242	accggtacggttaaaagcttctgtgacaaaacggcgttgggtggtgtccgtcgtgggtgcc	6301		
Db	5701	ACCGGTACGGTTTAAAGGCTTTGCTGACAAACCGCGGTGGTGGCGTCTCCGGTGGGTGCC	5760		
Qy	6302	tcagggtcgaacgcggtatcccaaatgtgcagcaggttgaggtcagtgccaccagcaccc	6361		
Db	5761	TCAGGTGCTGAACCCCGTATCCAAAATGTGGCAGCAGGTGAGGTCAAGTCCACCAAGCAC	5820		
Qy	6362	gatcgcgtcaatggtgaccagttgtacaaagccaccccaagcatttgcacaacgcaacaa	6421		
Db	5821	GATCGGTCAATGTTAGTAGCTGTTGTAACAAAGCCACCAAGCAATTGCCAACGCAACCAAT	5880		
Qy	6422	gagcttgaccatcgtatccaccacaaacgaaataaagccaatcagggatttccatcagcg	6481		
Db	5881	GAGCTTGACCATCGTATCCCAACCAACGAAATTAAGGCCAATCGAGGATTTTCATCAGCG	5940		
Qy	6482	atggcagtggcgtccatgccaagcctacattccttggcagatccatggttaccgggggt	6541		

Db	5941	ATGGCGATGGCGTCCATGCCACAAGCCTACTCTCTGGCAGATCCATGTTACCGGGGT	6000		
Qy	6542	attgcacacacaacaggtcaaggtcggtgaggtggtgactgtcgaagctgtcgggataat	6601		
Db	6001	ATTGCCACCCACAACCGGTCAAGGTGCGGTGGCAGTGGGACTGTCTGAAGCTGTCTGGATAAT	6060		
Qy	6602	ggtcaatgggtatttaaaatcaatgggttccagccgatacccaaggcccatgtaggggcccga	6661		
Db	6061	GGTCAATGGGTATTTAAATCAATGTTTACGCCGATACCCAGGCCATGTAGGGGGCGCA	6120		
Qy	6662	gttgggtgcaggttttcaacttt	6682		
Db	6121	GTGGGTGCAGGTTTTCACATT	6141		
RESULT 6					
AX079914	AX079914	5976 bp	DNA	PAT 22-FEB-2001	
LOCUS	Sequence 2 from Patent WO0107619.				
DEFINITION					
ACCESSION	AX079914				
VERSION	AX079914.1	GI:13159437			
KEYWORDS					
SOURCE	Moraxella catarrhalis.				
ORGANISM	Moraxella catarrhalis				
REFERENCE	1 (bases 1 to 5976)				
AUTHORS	Loomsore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.				
TITLE	Recombinant high molecular weight major outer membrane protein of moraxella				
JOURNAL	Patent: WO 0107619-A 2 01-FEB-2001;				
CONNAUGHT LABORATORIES LIMITED (CA)					
FEATURES	Location/Qualifiers				
source	1..5976				
/organism="Moraxella catarrhalis"					
/db xref="taxon:480"					
BASE COUNT	1972 a 1360 c 1355 g 1289 t				
ORIGIN					
Query Match	85.7%;	Score 5976;	DB 10;	Length 5976;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 5976;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy	707	gtgacgtgtgaacgctcagtcagtcagtcgtcttatctcaaaaaaagataccaaacatc	766		
Db	1	GTGATCGGTGCAACGCTCAGTGGCAGTCTTATGCTCAAAAAAAGATACCAACATATC	60		
Qy	767	gcaattggtgaacaaacccagcagcgtcagcagcgtcagcagcgtcagcagcgtcagcag	826		
Db	61	GCAATTGGTGAACAAACCCAGCAGACGCTCAGGCACCTGCCAAGGGCGAGGTGATCGA	120		
Qy	827	gccattgctattgttgaaaaatgctaaccacagggcggtcaagccatcgccatcggtagt	886		
Db	121	GCCATTGCTATTGTGTAATGCTAACGCACAGGGCGGTCAAGCCATCGCATCGGTAGT	180		
Qy	887	agtaataaacgtcgaatgggaagcagtttggataagataggtaccgatgctacgggtc	946		
Db	181	AGTATAAAGTCTCAATGGAGACGATTTGGATAGATAGGTACCGATGCTACGGGTCAA	240		
Qy	947	gggtccatcgccatcggttgggtgagttaaaagcgtagtggtgagtcgctcagtcac	1006		
Db	241	GAGTCCATCGCCATCGGTGGTGTATTAAGGGCTAGTGGTGTATGCTCGATTGCCATCGGT	300		
Qy	1007	agtgatgactacatttcttcttgcagcagtcagtcagtcagtcagtcagtcagtcagtcag	1066		
Db	301	AGTGATGACTTACATTGCTTGTGATCAGCATGTTATCTTAAACATCCGAAAGGTACTCTG	360		
Qy	1067	attaacgactttattaaacggccatcgatgatttaaagaataacgaagtcacgaagtc	1126		
Db	361	ATTACGACTTATTAAACGCCATGCAGTATTAAAGAAATACGAAGCTCAAGGATAT	420		







Db 2581 TTTAATGTTAACTAGTGTAGATGAAGATGCCCTTGTAAAGCCAAAGACATCGCCGAAAAAT 2640  
Qy 3347 ctaaacacctagcaaggaataatcacacacaaagccagacagacacacgcccctcaaa 3406  
Db 2641 CTAACACCCCTAGCCAGGAATTTACACCAACCAAGGACAGCAGACACCGCCCTCAAA 2700  
Qy 3407 acctttaccgtttaaaggttagatgaataataatgtctgatgcgccaacgcccataccc 3466  
Db 2701 ACCTTTACCCTTAAAGGTAGATGAATTAATTAATGCTGATGACGCCAAGCCCATCACC 2760  
Qy 3467 gtgggtcaaaagacgcaaaataatcaagtcacacccctcaacactcaaaagtgaaacaggt 3526  
Db 2761 GTGGGTCAAAAGAACCGCAATTAATCAAGTCAACACCCCTAACACTCAAAAGGTGAAACGGT 2820  
Qy 3527 ctttaatataaacgcaaaaatgggtacaggttacctttgacattaacacacacacagcgt 3586  
Db 2821 CTTAATATTAACCGCAAAATGGTACGGTTACCTTTGGCATTAACCCACCAAGCGGT 2880  
Qy 3587 cttaaagcggcaaaagcacccctaaacgacggtgctgtgtctattataaaacccactgggt 3646  
Db 2881 CTTAAAGCCGCAAAAGCACCTTAACGACGGTGGCTTGTCTATTAAAAACCCCACTGGT 2940  
Qy 3647 agcgaacaaaatccaaagtcgggtgctgagtcgaggttgccaaaggttaataataatgggt 3706  
Db 2941 AGCGAACAAATCCAAGTCGGTGCATGCGCTGAAGTTTGCCAAAGGTTAATAATAATGGT 3000  
Qy 3707 gttgtaggtgctggcattgatggcacactgcqattaccagagatgaattggctttact 3766  
Db 3001 GTTGTAGGTGCTGGCATTTGATGGCACAACCTCGCATTACCAGAGATGAATTTGGCTTTACT 3060  
Qy 3767 gggactaatggctcacttgataaaagcaaaacccacccctaaagcaagacggtcattaaogca 3826  
Db 3061 GGGACTAATGGCTACTTGTATTAAGCAAAACCCCACTTAAGCAAAAGACGGCATTAACGCA 3120  
Qy 3827 ggtgtaaaagattacaaacattcaatcagtgagatgctccaaagacagccatgatgct 3886  
Db 3121 GGTGTAAAAAGATTACCAACATTCATCAGGTGAGATTGGCCAAAACAGCCATGATGCT 3180  
Qy 3887 gtgacagcggcgaagattatgatttaaaacccgaacttgaaacaaataatcagcagttact 3946  
Db 3181 GTGACAGCGCGCAAGATTATGATTAAAAACCGAACCTTGAAAAACAAAATCAGCAGTACT 3240  
Qy 3947 gccaaacagacaaacactattacacgaattctcagttacacactcaaaagacctctgatctacc 4006  
Db 3241 GCCAAACAGCACAAATCTATTACAGCAATTTCTCAGTAGCAGATGAACAAGGTAATAAC 3300  
Qy 4007 ttacggttagtaacaccttactccagttatgacacctcaaaagacctctgatctacc 4066  
Db 3301 TTTACGGTTAGTAACCCCTTACTCCAGTTATGACACCTCAAGACCTCTGATGTCATCACC 3360  
Qy 4067 ttgacggtgaaacacggtcattaccaccaaaggtataaataaaggttggtgctggtggccatt 4126  
Db 3361 TTTGCAAGTGAACCGCATTAACCAAGGTAAATAAAGGTGTCGTCGTGGGCAT 3420  
Qy 4127 gccaaaccaaaggtttaacacgctaaagtcgaacgtgggttaataataatggcaaaagc 4186  
Db 3421 GACCAAAACAAAGGTTAAACACGCTAAGCTGACCGGTGGTAAATAATAATGGCAAGGC 3480  
Qy 4187 attgtcattgacagcgaataatgtaataatcattcacgactaagcaaacacttagct 4246  
Db 3481 ATTGTCAATTGACGCCAAATGGTCAAAATACCAATCAACAGGACTAAGCAACACTTAGCT 3540  
Qy 4247 aatgttaccatgataaaaggtgagctacgacacacagacaggggcaataataatcaaaagac 4306  
Db 3541 AATGTTACCAATGATAAAGGTAGCTAGCAGCACAGAACAGGGCAATATATATCAAGAC 3600  
Qy 4307 gaagcaaaacccgtgcgcgaagattgttgatgctgaagcagaggtttaacttgaa 4366  
Db 3601 GAAGCAAAACCCGTCGCCCGCAGCATGTTGATGTGCTAAGCGCAGGCTTTAACTTGCAA 3660  
Qy 4367 ggcataatggtagcgggttgactttgtctccacttatgacacgcgtcaactttgcccagatggc 4426  
Db 3661 GCCAATGTTGAAGCGGTGACATTGTTCTCCACTTATGACACCGTCAACTTTTGCCGATGTC 3720

Qy 4427 aatgcaccaccgctaaagtgacctatgatgacacaaagcaaaacccagtaaaagtggtctat 4486  
Db 3721 AATGCCACCACCCTAAGGTGACCTATGATGACACAAAGCAAAACCAAGTAAAGTGGTCTAT 3780  
Qy 4487 gatgctaattgatgatatacaacacattgaagttaaagataaaaaacttgctggttaaaac 4546  
Db 3781 GATGTCATATGATGATACAAACCTTTGAAGTTAAAGATAAAAAATGGCGTTAAAAACC 3840  
Qy 4547 accacattgaccagttactgacacaggtgctaaataatgtccctaaagcaaatcaagctact 4606  
Db 3841 ACCACATTGACCAAGTACTGGCACAGGTGCTTAATAAATTTGCCCTAAGCAATCAAGCTACT 3900  
Qy 4607 gggatgctgcttcaagggcagtgatgctgtgctctatcaaacacttatctctggtggac 4666  
Db 3901 GGGATGCGCTTGTCAAGGCCAGTGATGCTGCTCATCTAAACACCTTATCTGCGGAC 3960  
Qy 4667 atcaaacctgccaaaaggggcaagccaaagcaaacactcagcaggtctatggtatgctgat 4726  
Db 3961 ATCCAAATGCCAAAGGGCAAGCCAAAGCAAACTGGTGCCTCAAGCCCAACCCCA 4020  
Qy 4727 ggaataaggttcaatgatgacagttaccgaatacaagttactatcaagccaaaaatgatggc 4786  
Db 4021 GGCATTAAGTCTATCTATGACAGTACCGATACCAAGTACTATCAAGCCAAAATGATGGC 4080  
Qy 4787 acagttgataaaaccaaaggaagttgccaagacaaaactggtgcgccaagcccaaacccca 4846  
Db 4081 ACAGTTGATAAAACCAAGAAAGTTGCCAAAGCAAACTGGTGCCTCAAGCCCAACCCCA 4140  
Qy 4847 gatgacattgctcaaatgaatgtaaatcagttacttaacaaagacaaagtaaatgat 4906  
Db 4141 GATGGCATTGGCTCAAAATGAATGTCAATCAGTCTATTACAAAGAACAAATGAATGAT 4200  
Qy 4907 gccaaataaaagcaaggtcaatgaagacaaacgcttctgtaagagacttgaaagac 4966  
Db 4201 GCCAATTAAGCAAGGATCAATGAAGACAACGCCCTTTGTTAAAGGACTTTGAAAAAGCC 4260  
Qy 4967 gctctgataaaaccaaagacgcccaggttaactgaggtgaggtggtttaaattgcggtgccc 5026  
Db 4261 GCTTCTGATACAAACCAAAACGCCGAGTAACTGTTGGGTGATTTAAATGGCGTTGCC 4320  
Qy 5027 caaacacgctgacctttgacggggatcacagcgcaacgctcaaaacactggcgagact 5086  
Db 4321 CAACACCGCTGACCTTTGACAGGGATACAGCACACGCTTAAACACTGGCGGAGACT 4380  
Qy 5087 ttgaccatcaaaagtgggcaaacagacacccaaataagcttaaccgataataacatcgtggtg 5146  
Db 4381 TTGACCATCAAAAGGTGGCAAAACAGACACCAATAAGCTTAACCGATTAATAACATCGGTGTG 4440  
Qy 5147 gtagcaggtactgatggttcaactgtcaaaccttgccaaagacctaaccatcttaacagc 5206  
Db 4441 GTAGCAGGTACTGATGGCTTCACTGTCAAACTTGCCAAAGACCTTAACCAATCTTAACAGC 4500  
Qy 5207 gtaatgaggtggccacaaatgataacaaagcggtgctctttttagtactcaagcggt 5266  
Db 4501 GTTAATGCGAGTGGCACCAAAATGATGACAAAGCGGTCTTTTGTAGACTCAAGCGGT 4560  
Qy 5267 caagccaaagcaaacacccctgtctgaatgccaatgggctggacactgggtggcgaagtc 5326  
Db 4561 CAAGCCAAAGCAAAACACCCCTGCTGCTAAGTGCCTAAGTGCCTGGACCTGGGTGGCAAGTC 4620  
Qy 5327 atcagtaattgggcaagggcacaaagataccagcgtgcctaattgatacaacagttaaac 5386  
Db 4621 ATCAGTAATGTTGGGCAAGGCAAAAGATACCGACGTGCCAATGTCATACACAGTTAANC 4680  
Qy 5387 gaagtacgcaactgtgtggttctgtgtaattgctgtaataacacgctgaacgcaatcag 5446  
Db 4681 GAAGTACCAACTTGTGGTCTTGGTAACTGCTGTAATGATGATAACGCTGACGGCAATCAG 4740  
Qy 5447 gtaaacattgcacacatacaaaaagacccaaattcaggttcatcatctaaacgcaactgctc 5506  
Db 4741 GTAAACATTGGCGACATCAAAAAGACCCAAATTCAGGTTTATCATCTAACCGCACTGTC 4800



Db 42616 TAAGCTAACCCACCATGAATCACATCTATAAGTCAATCTTTAAACAAAGCCACAGGCAT 42675  
 QY 588 ttatggcagtgacagtacgcgaatccccacagcacgagggggg-tagctgtgtaca 646  
 Db 42676 TTATGCGCGTGGCGAATATGCCAATCCACAGCACGGGGGGATAGCTGTCTACA 42735  
 QY 647 gggcaagtggcagtgatgactgactgagcttgcgcgtatgcccgtcgtgctc 706  
 Db 42736 GGGCAAGTTGGCAGTGTATGCATCTGAGCTTGGCCGTATTTGGCGGCTCGCTGCTC 42795  
 QY 707 gfgatcgatgaacgctcagtgagtgatgctgctcaaaaaaagatacacaacatc 766  
 Db 42796 GTGATCGGTGCCACGCTCAGTGGCACTGCTTATGCTC-----AAGGTATC 42840  
 QY 767 gcaattggtgaacaaacacagcaagacgctcagggcactgcaagggcgagcggtatoga 826  
 Db 42841 GCAATTGGTCAACAAACAA---CAACTCAAAAGGCGATGCCGAGCGACAGGTAAATCAA 42897  
 QY 827 gccattgctattggtgaaatgctaacgcacagggcgctcaagccatgccatcggtagt 886  
 Db 42898 TCCGTGTCTTGGTGAAAGACTGTTGCAAGTGGCAGTCAAGCCATGCCATCGTGTGT 42957  
 QY 887 agtaataaaactgtaagtgaagcgcttggataagataggtaccgatgctacgggtcaa 946  
 Db 42958 AGCCAAA-----CAAAACAATAATAGTAAAGTATCAGGTGCCAACGCCAAAGGTGAA 43011  
 QY 947 gagtccatgcacatcggtggtgtagtaaaagcgctagtgatgctcgtatgcatcggt 1006  
 Db 43012 GAGTCCATGCCATCGGTGGTGTATTTGGCTCAGGTCCTCGCTCGATTTGCCATCGGT 43071  
 QY 1007 agtgatgacttacattgctgacgatggtatccataacatccgaaggtactctg 1066  
 Db 43072 AGTGATGACTTATTT-----GGAGAGTAAAGGCTTGTATACACTCCACGAGAG 43122  
 QY 1067 attacgatcttattaacggccatgcagttattaaagaataacgaagctcaa---gagat 1123  
 Db 43123 ATTAAAAAATTTAAACAAATTCAGAAATTAACAGAAATACGAATACTGATCCCAAT 43182  
 QY 1124 aatgattgaaatataagacgaacacgcaagcagcagcagcagcagcagcagcagcagc 1183  
 Db 43183 AGTAATCAAAATTTATACACAAAGCAAGCAAGGACACCCAGTATTCAGTGGGAACC 43242  
 QY 1184 atgtcatgacaggtgctattttccacgctttggttacaggggcaacagctcaaaagt 1243  
 Db 43243 ATGGCAGGCAGAGGTCATTTTGCCACGCTTTTGGTACAGCGCAATAGCTAAAGC 43302  
 QY 1244 gctattcttggcagtggtcttgcgcacagcagcagcagcagcagcagcagcagcagc 1303  
 Db 43303 AACTATTCTTGGCAGTGGTCTTACCGCAAGCGGACACAGAAATCTTCAATCGCTCT 43362  
 QY 1304 ggtctgatgaacatctagctcgttggagcgatagcccttgggtcaggtactcgtgct 1363  
 Db 43363 GGTCTAATGCACAGCTAGCAAGTTTGCAGCGATAGCCATTGGTGGAAATCTAAATCT 43422  
 QY 1364 cagctacagggcagttatgcccctagtgtaaggtcttctgtcactcagagtgataaat 1423  
 Db 43423 GAGTTGGGTGGAGGTGTGGCTTAGTTTAAATTTTACATGATCTGCTAATGATAATAT 43482  
 QY 1424 tctagacggtcctatacacaaataccagggcactagaccccaagtttcaagccaccaa- 1482  
 Db 43483 CTAAATGGACCTTATAACACAGCGGTGAACATTTACACAGGATCATGAGGCCACCAA 43542  
 QY 1483 -----taatacgaaggcggttccactttccatttggttagtaactctatacaacgt 1531  
 Db 43543 AACGGTCAGCGGAAGATACAGTGGATATATTTTCCATTGGTAAATAGTATCTCAGACGT 43602  
 QY 1532 aaaaatcatatgctggtgaggtgtaataaaacagcagcgtcgtcaatggtgcacagcta 1591  
 Db 43603 AGAATCATTCATGTCGGTGGAGGTTGTATGGCCACCGCTGCTCATGTGGCACAGCTA 43662  
 QY 1592 gaagcgtggtggaagtggtgctaaagcagctagataattacttttcaggggtgataacagt 1651

Db 43663 GAAGCGGTGCTGAACGTGGCTAAGAACGCTAAATTTACTTTTTAAGGGTGATG---ACAGT 43719  
 QY 1652 actgacgtaaaaataggtttggaataacttaactattaaagtg---gtgcagagacc 1708  
 Db 43720 ACTGGCGTCAAAAAGGTTTCGACGAGGATTAACTATTAAAGCGGGGAGCAATGCC 43779  
 QY 1709 aagcattaccacgataataataatcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1765  
 Db 43780 GGTGAATTAACGACATAAATCATCGGTGTGTAAAGATAATGATATGCTGGTCTG 43839  
 QY 1766 aaggttaacttgcataaacttttaacaacttctactgaggtgaatacaactataaat 1825  
 Db 43840 AAGTTTAACTTGTCTAAAAATTTAACTGATCTTGAACAGTTAAACACGCAAAACCTTAAC 43899  
 QY 1826 gccacaaccacagtaagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1882  
 Db 43900 GCCACAGGTGCAATCACCGTAAATAATAATAGTACTACCAATACGGCTACATATTATGGCT 43959  
 QY 1883 gatagtttaaccttaccacgcccacatacacagggcagtcacagcacaagcaaaacgctctat 1942  
 Db 43960 GGTGGTTTAACTGTTAAAGATATTTCATA---ACAACAAATCCCAAGTCAGTGTG---AT 44013  
 QY 1943 ggcgttaataaggggtgaagtttacttaataataatgcagaacacacagcagcaatcgccactact 2002  
 Db 44014 GGCATTTAAATTTACTGATATGAATGGTGGTGTACATCAGCCCCAGTAGCAGCGCGCACTACT 44073  
 QY 2003 cgtattaccagagataaaattggcttggctcgagatggtgtagtggatgataaaacaaagca 2062  
 Db 44074 CGTATTACCAGAGATACAAATTTGGCTTTGCCAAGAGAGTGTGTACAGTTGATGACAATTC 44133  
 QY 2063 ccatattgggtataaaacaaacttaagtggttagtggttagtggttagtggttagtggttagtggt 2122  
 Db 44134 CCATATTGGATACGCAAAAGCTAAAGTTGGCAATTTGAGATTAAATAAGACACACTGGT 44193  
 QY 2123 attgatgcagtaataaaagatcagtaattcttgcgaaggtagcagtgctaaagcagtcg 2182  
 Db 44194 ATTATGTCAGGTATCAAAAGATTACCGACTTACTGTGTCGACGAGCATGCCGATGCT 4253  
 QY 2183 gttaccatcgaaacagctcaaaagccgacgactctttaaagcagggcgctggcactcag 2242  
 Db 44254 GTTACCATCAACAGCTCAAAAAGCCCAAGCTTAATTTAACCCCAAGCAATGGCATTTGAG 44313  
 QY 2243 gtcacacctactgaaa---tatcagttgatgtaagagtggaatgttaccgcccacact 2299  
 Db 44314 ATTACAAATACAAACAGTCTATTAGTTGCTGCTAATGTTATGTTACCCCCCATCT 44373  
 QY 2300 tacaacattggcgtgaaacccacgagcttaacagtgatgagtggaatgttaccgcccacact 2356  
 Db 44374 TATACCATTTGGGTGAAACCAACCACTCTGTACCAATAATGGCCCTAGTGTAGTAAATTT 44433  
 QY 2357 agtgtaagggtagtggtagc---aacaatagcttagttaccgcccgaacatttggcaagc 2413  
 Db 44434 GCTGTTAAGGATAGTAATACGCTAATACGCTGACAGTGTCTGTACAAAGCTTTACCGTTAAAGT 44553  
 QY 2414 tatctaaatgaagtcacacgagcgtgacagtgctctacaaagcttaccggttaaaagaa 2473  
 Db 44494 TATCTTAATGAAGTCAATCGAACGCTGACAGTGTCTGTACAAAGCTTTACCGTTAAAGT 44553  
 QY 2474 gaagacgatgtagcagcgaacgctatccagctggggtgaagatacagcaaaaaatgccgcg 2533  
 Db 44554 GGTGATGTCAGCAAGCAACACAGCATCAGCTAGATATAAAACCAACAGCAACAGTCTAACCCGTA 44613  
 QY 2534 gcagtcagcatcttaaaactcaaaagtaaaaaacggttctaaacggttggctaccacaaaaaagat 2593  
 Db 44614 GCAGGGGAT-----AACGGCATTTACTGTTAAACGGATACAACTACGAAAAA 44661  
 QY 2594 ggtacggtttacgtttggcttagccaagatagcgtctgacatttggcaaaagccacctta 2653  
 Db 44662 GTCACGTTGGCATTC-----ACCAAGCAATGGCTTAACCAACGCTTAAGCTGACCGTG 44715  
 QY 2654 aacaacgatggtctgactgttaaaatccacgaacaaatccaagtcggtgcttaagtc 2713  
 Db 44716 GGTAGCAATGAATAAATCAATCAATTTGTTATTTGACCAATGGT----- 44756

Qy	2714	attaaattactaatgtgaatggttagttaatccaggctactggcattgcaaataccgcgtgcg	2773
Db	44757	-----	44756
Qy	2774	attaccagagataaaattggcttgcgtggttctgtgatggtgcagttgatacaaacaaacct	2833
Db	44757	-----	44756
Qy	2834	tatcttgatcaagacaagctacaaagttggcaattgaagattaccaaacactggcattaac	2893
Db	44757	-----TATCGGTAAC	44766
Qy	2894	gcaggtgtaagccatcaccaaggctgccccaacactgctagcattgcga---tcaa	2950
Db	44767	GACCCAAAACATCATCAAAAGGCTGCTCAACCTGACTGACATTACCAACACACAGAT	44826
Qy	2951	agtgcgcgaacatagaaactgggcaatacaatccaaagacaagaacaaatccaaacgtgc	3010
Db	44827	AATACACACACCAGAACAGACAATGACGCCCAAGGCAAG--AATCCAACGCTGCC	44883
Qy	3011	agcattaatgatattataatcacaggctttaacctcaaaaaataatacaaacccccattgac	3070
Db	44884	AGTATCAAAAGTGTCTAAACGCAAGGCTTTAACTTGAAGGCAATGGCAAGCCGTTGAC	44943
Qy	3071	tttgtctccacttatgacattgttgacttgcgaatggcaatggcaatgcacacgcgcacagta	3130
Db	44944	TTGTGCTCTACTTATGATACCGCTCAACTTTGCCAATGGCGATGGCACCACTGCCACAGTT	45003
Qy	3131	acc-----catgatcacgcgttaacaaacacagtaagtggtatatgtatgtgaattgtggt	3184
Db	45004	ACCTATATATTATTAATGAAGGCAAAACCAACGACAGACGTGGTCTATATGTCATATGGAT	45063
Qy	3185	gatacaaccttcatctaaacaggcactgatgacaataaaaaacttggcgtcaaaaacacc	3244
Db	45064	GATACACCAATTCTCTGTGGCAAAATAAATGGTATAAATAAACTTAGCGTAAAAACCAAC	45123
Qy	3245	aaactgaacaaaacaaagtctaatggttaatacacgaactaaacttaattgttaactctagt	3304
Db	45124	ACATTGACCAAAACAAATGGTGCAAAATGGTAAATGCACATAAATTTAGTGCCAAC-----	45177
Qy	3305	gatgaatgcctctgttaaagcgaagaagacatcgccgaaaaatctaaacacactgaccaag	3364
Db	45178	AATGGCATGCCCTTGTTTAAGCGCAAAAGGTATTGCCGCAACATCTAAACATCTAGCCGAG	45237
Qy	3365	gaaattcaacaccacaagaagcaagacagacacgcgcctacaacactttacogttaaanaag	3424
Db	45238	GAAATTCACACCCCAAGGCACAGACAGACACCGCCCTACAAACCTTTAAAGTCAAAAA	45297
Qy	3425	gtagatgaataataatgctgtatgagccaaagccatcacctgggttcaaaaagaagca	3484
Db	45298	GACGGTGCAACCAATGATGATG-----ACACCATCACCGTGGGT-----	45336
Qy	3485	aataatcaagtcaacacacctcaactcaaaaggttgaaacggtctttaattataaacacgac	3544
Db	45337	AAAATGCGGTGACACCCCTAACATTCAAAAGGTGAAAACAGGCTCTAACGGTTGCTACCAA	45396
Qy	3545	aaaatggttcggttacctttggcattaaacacacaagcgtcttaaaagccgg---caaa	3601
Db	45397	AAAGATGTTACGGTTTACCTTTTGGGATTTAAACACCCAAAGCGGTCTTTAAAGCGCGCAAC	45456
Qy	3602	agcaccttaaacgacggttggtcttctattaaaacccccactgggtgagcaacaataccaa	3661
Db	45457	ACCACCTTAACAAGATGGCTTGCTATTAAAANCCCCACTGGTAAAGACAATCCAA	45516
Qy	3662	gtcgggtcgtggtggtgaagtttgcaaggttaataataatgggtttaggtgctggc	3721
Db	45517	GTCGGTGTGTAGCGGTGAAGTTTGCCAAGGTTTAATAATAATGCTGTATAGTGTCTGGC	45576
Qy	3722	attaatggcaactcgcattaccagagatgaattggctttactgggactaataggctca	3781
Db	45577	ATTGATGGCAACGCCGTATCCACCAAGATAAAATGCTTTGTGTGGGGCTTAATGGGTCA	45636

Qy	3782	cttgataaaagcaaacccaccactaagcaaaagcggcattaaacgaggtggtgtaaaagatt	3841
Db	45637	CTTGATACCAACCAAAACCCACCTTAAGCAAAAGACGCATTAACGCGAGGTGTTAAAAAGATT	45696
Qy	3842	accaacattcaatcagtgagattgccaaacagccatgatgctgtgacaggcggaag	3901
Db	45697	ACCAACATTCAATCAGGTGAGATTGCCAAAACACCCATGATGCTGTGACAGCGGTCAG	45756
Qy	3902	attatgatttaaaacccgaacctgaaacaaaaatcagcagtlactgccccaaacagcacaa	3961
Db	45757	ATTATTATGATTAGAAACCGAACTTAAGAAACAAAATCAGCAGTACTGCGCAAAACAGCACAA	45816
Qy	3962	aactcattacacgaattctcagtagcagatgaacaaggtataaacctttacggttagtaac	4021
Db	45817	AACTCATTTACAGCAATCTCAGTAGCAGATGAACCGGTAATAACTTTACGGTTAGTAAC	45876
Qy	4022	ccctactccagttatgacacctcaagacctctgatgtcatcacctttgcaggtgaaac	4081
Db	45877	CCTTTACTCGAGTTATGACACCTCAAGACCTCTGATGTCTACCTTTGCAGGTGAAAAC	45936
Qy	4082	ggcattaccaccaagttaataaaagtggtgctgctggcgcatlgaccaaaccaaagcg	4141
Db	45937	GSCATTACCAACCAAGSTAAATAAAGGTGTGGTGCCTGTGGCATTGACCAACCAAGGC	45996
Qy	4142	ttaacacgcctaaagctgacogtgggttaataataatggcgaaggcatttgttcattgacagc	4201
Db	45997	TTTAACACGCGCTTAAGCTGACCGTGGGTGATATAATAATGGAAGGCAATGTCTATTGACAGT	46056
Qy	4202	caaaatgtcaaaataccatcacaggaactaagcaactctagcttaactgttaacttaacatgat	4261
Db	46057	AAAGATGGTCAAAATACCATCACGACTAAGCAACACTCTAGCTTAATGTTTACCAATGA-	46115
Qy	4262	aaaggtagctgacgcaccacagaaacagggcaataaatacaagacgaagacaaacccgt	4321
Db	46116	-----TGGTGACGGACACGCCTAAGCCCAAGGCTTGGCGATACACCGACAAAACCCGT	46170
Qy	4322	gcccgcagcattgttgatgtgctgaagcgcagcgctttaacttgcaaggcaatgggtgaagcg	4381
Db	46171	GCTGCCAGCATTTGGTGATGTATTAAATGCGAGCGTTTAACTTGAAGGCAATGGTACAGCG	46230
Qy	4382	gttgacttgtctccacttatgacacctcaacttgcgattgcaatggcaatgccaccacgct	4441
Db	46231	GTGTACTTGTCTCCACTTATGACACTGTCAACTTTGGCGATGGCAATGCCACCACCGCT	46290
Qy	4442	aaagtgacctatgatgacaaagcaaaacccagtaaaagtggtctatgctgctcaattggat	4501
Db	46291	AAGGTGACCTATGATGACAGGACAAAACCCAGTCAAGTGGTCTATGATGTCAATGTGGAT	46350
Qy	4502	gatacaaccattgaagttaaagataaaaaacttggcgtaaaaaaccacacattgaccagt	4561
Db	46351	GATACACCACTTGAAGTTAAAGATAAAAAACTTTGGCGTAAAAACCCACACATTTGACCAGT	46410
Qy	4562	actggcacaggtgctaataaaattgcccctaagcaatcaagctactggcgatgcgcttttc	4621
Db	46411	ACTGGCAGAGTGCTTAATAAATTTGCCCTAAGCAATCAAGCTACTTGGCGATGCGGCTGTC	46470
Qy	4622	aaaggcagtgatatcgttgctcatctaaacaccttatctggcgacatccaaactgccaaa	4681
Db	46471	AAGGCCAGTGATATCGTGCTCATCTAACACCTTATCTGGCGACATCCAAACTGCCAAA	46530
Qy	4682	ggggcagccaaagcgaacactcagcggtatgtggatgctgctggcaataaggtcatc	4741
Db	46531	GGGGCAAGCGAAGCAACACTCAGCAGGCTATGTGGATGCTGTATGTCGAATAAGGTCATC	46590
Qy	4742	tatgacagtaccgataaacagttactatcaagccaaaaatgatggcacagttgataaaacc	4801
Db	46591	TATGACAGTACCGGATTAACAAGTACTATCAAGCCAAAAATGATGGCACAGTTGATAAAAACC	46650
Qy	4802	aaagaagtggccaaagacaaactggtgcccaagcccaaacccagatggcacattggct	4861
Db	46651	AAAGAAGTTGCCAAGACAAACTGTGTGCCCCNAGCCCCAAACCCCAAGATGGCACATTGGCT	46710
Qy	4862	caaatgaatgtcaaatcagttcattacaagaagcaacgaagtaaatgatgccaatcaaaagcaa	4921

Db 46711 CAAATGAATGTCAATAGTCAATTAACAAAGCAAGTAATGATGCCAATAAAGCAAA 46770  
Qy |||||  
Db 4922 ggcataatgaagcaaacgcttctttaaagagcttgaagagcgtcttctgataacaaa 4981  
Qy |||||  
Db 46771 GGCATCAATGAAGCAACAGCCTTTGTTAAAGGACTTGAAAAGCGCTTCTGTATAACAAA 46830  
Qy |||||  
Db 4982 accaaaaacgcgcagtaactcgtggtgatttaaatgcggttgcccaaaacgcgtgaac 5041  
Qy |||||  
Db 46831 ACCAAAAACGGCGCAGTACTGTGGGTGATTAAATGCGGTGGCCAAACACCGCTGACC 46890  
Qy |||||  
Db 5042 ttgcagggatatacaggcacacgcgttaaaaaactggcgagagctttgaccatcaaaagt 5101  
Qy |||||  
Db 46891 TTTCAGGGGATACAGGCACAAACGGCTAAAAAATGGCGGAGACTTTGACCATCAAAAGT 46950  
Qy |||||  
Db 5102 gggcaaacagacacacaaataagctaacccgataataacatcgtgtggtgtagcaggtaactgat 5161  
Qy |||||  
Db 46951 GGGCAACAGACACCAATAAAGCTAAACCGATAATACATCGGTGGTGAAGCAAGTACTGAT 47010  
Qy |||||  
Db 5162 ggttcactgtcaaaacttgcaaacacacctaaccatcttaacagcgttaatgcaggtggc 5221  
Qy |||||  
Db 47011 GGTCTACGTCTCAAACTTTGCCAAGACCTTACCAATCTTACAGCGTTAATGCAGGTGGC 47070  
Qy |||||  
Db 5222 accaaaattgatgacaaaggcgtgtcttctttagactcaagcgttcaagcccaagcaaac 5281  
Qy |||||  
Db 47071 ACCAAATTTGATGACAAAGGCGCTGTCTTTTGTAGACTCAAGCGGTCAAGCCAAAGCAAAAT 47130  
Qy |||||  
Db 5282 accctgttactaagtgcgaatggctggacctgggtgcaaggtcatcagtaagtggc 5341  
Qy |||||  
Db 47131 ACCCTGTCTTAAGTGGCAATGGGTGGACCTGGGTGGCAAGGTCTATCAGCAATGTGGGT 47190  
Qy |||||  
Db 5342 aaagcacaaaaatgacagcgtcgcgaatgtacacagtttaaacgaagtacgcgaacttg 5401  
Qy |||||  
Db 47191 AAAGGCACAAAGACACCGACGCTGCCAATGTACAACAGTTAAACGAAGTACGCAACTTG 47250  
Qy |||||  
Db 5402 ttgggtcttggtaagtctgttaatacagctgcagcgcaatcaggttaaacattgccgac 5461  
Qy |||||  
Db 47251 TTGGGTCTTTGGTAATGCTGGTAATGATAACGCTGACGGCAATCAGGTAAACATTGCCGAC 47310  
Qy |||||  
Db 5462 atcaaaaaagaccnaatlcaggttcatctataacgcgactgtcatcaaacagcagcagc 5521  
Qy |||||  
Db 47311 ATCAAAAAGACCCCAATTCAGGTCTCATCTATACCGGCACTGTCAATCAAGCAGCGCAG 47370  
Qy |||||  
Db 5522 gtacttggcgttaaaagtataacagataccgaaaaacttgccactgtggtatacaagt 5581  
Qy |||||  
Db 47371 GTACTTGGCGGTAAAGGTAAATACGATACCGAAAAAATTTGCCACTGGTGGTATACAAGTG 47430  
Qy |||||  
Db 5582 ggcgtgataaagacgcaacgctaacgctgatttaagcaatgttgggtcaaaacccaa 5641  
Qy |||||  
Db 47431 GGCCTGGATAAAGACCGCAACGCTAACGGCGATTTPAAGCAATGTTTGGGTCAAAACCCAA 47490  
Qy |||||  
Db 5642 aaagatggcagcaaaaaagccctgctgcgaacttaaacgcgcaggttcagaccactat 5701  
Qy |||||  
Db 47491 AAAGATGGCAGCAAAAAGCCCTGCTCGCCACTTATAACGCCGCGAGGTGAGACCAACTAT 47550  
Qy |||||  
Db 5702 ttgacaaacaaacccgcagaaagccattgacagataaatagaacaggtatccgtctcttc 5761  
Qy |||||  
Db 47551 TTGACCAACAAACCCCGCAGAAAGCCATTGACAGAAATTAATGAACAAAGGTATCCGCTCTTTC 47610  
Qy |||||  
Db 5762 catgtcaacgatggcaatcaacagcctgtgttacaaaggcgtaacggcattgactcaagt 5821  
Qy |||||  
Db 47611 CATGTCAACGATGGCAATCAAGAGCCTGTGTACAAAGGGCGTTAACGGCAATTTGATTTCAAGT 47670  
Qy |||||  
Db 5822 gacctgaagcactcagtggcgataagtttccagggcaaggcagataggtgaagccgc 5881  
Qy |||||  
Db 47671 GCTGCTGGCAAGCAATTCAGTGGCGATAGTTTCCAGGCCAAGGCAGATGGTGAAGCCGCC 47730  
Qy |||||  
Db 5882 gttgccatagcagcaaaaaccccaagcagcgaacaaatccatccctcgcctcgtgtgataacgca 5941  
Qy |||||  
Db 47731 GTTCCCATAGGCACACAAACCCRAGCAGGAGCGCAATCCATCGCCATCGGTGATAAGCA 47790  
Qy |||||  
Db 5942 caagccacggcgatcaatccatccatcggtacagggcaatgtgtagcaggtgaagcac 6001  
Qy |||||

Db 47791 CAAAGCCACGGCGGATCAATCTCATCGCCATCGGTACAGGCAATGTGGTAGCAGGTAAGCAC 47850  
Qy |||||  
Db 6002 tctgggtccatcgcgcagaccccaagcactgttaagctgtataacagttacagtgctgggtaaat 6061  
Qy |||||  
Db 47851 TCTGGGTGCCATCGCGCACCCAGCACTGTTAAGGCTGATAACAGTTACAGTGTGGGTAAT 47910  
Qy |||||  
Db 6062 aacaacagtttaccgatgcacactcaaacgatgtcttcttgggtgtggtggcaataaacaac 6121  
Qy |||||  
Db 47911 AACAAACAGTTTACCAGTGCCTAGACCGATGTCTTTGGTGTGGCAATACATCAC 47970  
Qy |||||  
Db 6122 gtaccgaaagtaactcgtgtgcttaggttcaaaccttgcacatcagtcagtcagcacacac 6181  
Qy |||||  
Db 47971 GTGACCAAAAGTAACTCGGTTCCTTAGGTTCAAACTCTGCCATCAGTGCAGGCACACAC 48030  
Qy |||||  
Db 6182 gcaggcacacaaccccaaaaaactctgcagcgcagcaggtacaaacccacacagcagtgca 6241  
Qy |||||  
Db 48031 GCAGGCACACAGCCCAAAAATCTGACGGCACAGCAGGTACAACACACAGCAGCGTGC 48090  
Qy |||||  
Db 6242 accggtacggttaaaagcttcttgacaaaacgcgcgcgtgtggtcgtctcctcgtgggtgc 6301  
Qy |||||  
Db 48091 ACAGGTACGGTTAAAGCTTTGCTGGACAAACGGCGGTGGTGGCGTCTCCGTGGGTGCC 48150  
Qy |||||  
Db 6302 tcaggtactcaacgcctatccaaaaatgtgcagcaggtgaggtcagtcagccacacagcacc 6361  
Qy |||||  
Db 48151 TCAGGTGCTGAACCCCGTATCCAAAAATGTGCAGCAGTCAGTFCAGTCCACACAGCACC 48210  
Qy |||||  
Db 6362 gatcggtcaatggttagcaggtgtacaaagcacaaccccaagcattgcacaacgcaacaaat 6421  
Qy |||||  
Db 48211 GATCGGTCAATGTGTAGCCAGTTGTACAAAGCCACCCCAAGCATTTGCCAACGCAACCAAT 48270  
Qy |||||  
Db 6422 gagttgacctcgtatccaccaaaacgaaaaaagcgaatgaagcgaatgttcacatcagcg 6481  
Qy |||||  
Db 48271 GAGCTTGACCATCTGTATCCACCAAAACGMAAATAAAGCCCAATGCAGGGATTTTCATCAGCG 48330  
Qy |||||  
Db 6482 atggcgatggcgctccatgccacaagcctacatctcctggcagatccatcgttaccgggggt 6541  
Qy |||||  
Db 48331 ATGGGATGGGTCCATGCCACAAGCCTACATTCCTGGCAGATCCATGGTTACCGGGGT 48390  
Qy |||||  
Db 6542 attgccaccaacacaggttcaaggtgcggtaggcagtgagcagtcgaagcgtcggataaat 6601  
Qy |||||  
Db 48391 ATTGCCACCCACACGGTCAAGGTGCGGTGGCAGTGGGACTGTCCAGCTGTCCAGCTGTCCGATTAAT 48450  
Qy |||||  
Db 6602 ggtcaatgggtatttaaatcaatgggttcagcgcgatacccaagcccatgtaggggcgcgca 6661  
Qy |||||  
Db 48451 GGTCAATGGGTATTAAATCAATGTTTCAGCGCATACCCAAAGCCATGTAGGGCGACGA 48510  
Qy |||||  
Db 6662 gttggtcaggttttccacttttaagccataaaatcgcgaagattttacttaaaaaatcaatct 6721  
Qy |||||  
Db 48511 GTTGTGCAGGTTCCTACCTTTTAAGCCATAAATCGCAAGATTTTACTTTAAAAATCAATCT 48570  
Qy |||||  
Db 6722 caccatgtgtataaaacagcagcagcagtcagtcataattactgatgctgagtttttta 6781  
Qy |||||  
Db 48571 CACCATAGTTGTATAAACACGATCAGCATCAGTCATATTACTGATGCTGATGTTTTTA 48630  
Qy |||||  
Db 6782 tcacttaaacatttaccgtcaagtgattctcttccacatgacccaaaacgcgcattga 6841  
Qy |||||  
Db 48631 TCACTTAAACCAATTTACCGCTCAAGTGTCTCTTCCACATGACCAAAATCGCCATTTGA 48690  
Qy |||||  
Db 6842 tcaataggtaaaccttattagtaaaatttatcaatgtagtgttagatatgggttaaaaaattg 6901  
Qy |||||  
Db 48691 TCATAGGTAAACTTATTTAGTAAATTTTATCAATGCTAGTTGTTAGTATGGTTAAAAATTG 48750  
Qy |||||  
Db 6902 tggcattgacccaaaaatgaccgatttatcccgaaaaattctcgtatgatcgctgacc 6961  
Qy |||||  
Db 48751 TGCCATTGACCAAAAAATGACCGATTTATCCGAAAAATTTCTGATTATGATCAGCTTTTCA 48810  
Qy |||||  
Db 6962 tgcaggtc 6969  
Qy |||||  
Db 48811 TAAATTC 48818  
Qy |||||

LOCUS	AX079920	6159 bp	DNA	PAT	22-FEB-2001			
DEFINITION	Sequence 8 from Patent WO0107619.							
ACCESSION	AX079920							
VERSION	AX079920.1	GI:13159440						
KEYWORDS								
SOURCE	Moraxella catarrhalis.							
ORGANISM	Moraxella catarrhalis							
	Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;							
	Moraxella.							
REFERENCE	1. (bases 1 to 6159)							
AUTHORS	Loosmore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.							
TITLE	Recombinant high molecular weight major outer membrane protein of moraxella							
JOURNAL	Patent: WO 0107619-A 8 01-FEB-2001;							
CONNAUGHT LABORATORIES LIMITED (CA)								
FEATURES	Location/Qualifiers							
source	1. .6159							
	/organism="Moraxella catarrhalis"							
	/db_xref="taxon:480"							
BASE COUNT	2035 a 1386 c 1385 g							
ORIGIN								
Query Match	58.8%;	Score	4101.6;	DB 10;	Length 6159;			
Best Local Similarity	81.3%;	Pred. No. 0;						
Matches 5063;	Conservative	0;	Mismatches 1009;	Indels 156;	Gaps 2			
Qy	542	atgaatcacatctataagtcattctttaacaaagccacagggacacatttatgacagtggca	601					
Db	1	ATGAATCACATCTATAAAGTCATCTTTAACAAAGCCACAGGCACATTTATGCCGTGGCG	60					
Qy	602	gagtcagccaaatcccacagcaagggggggtgctgtgtctacagggcaagtgtggcagt	661					
Db	61	GAATATGCCAAATCCCACAGTACGGGGGGGTAGTGTGCTACAGGGCAAGTTGGCACT	120					
Qy	662	gtatgcactctgagcttggccgtattggcgcgcctgcgtctctctgcgtcgtgcagcg	721					
Db	121	GTACGGCACTCTTAAGCTTTGGCCCGTATTGCCCGCTCGCTGCTCGTGTATCGTGGCAG	180					
Qy	722	ctcagtggcagtgcttatgctcaaaaagaagataccaaacacatctcaattgtagaaca	781					
Db	181	CTCAATGGCAGTGGCTATGCTCAACAAATTTACTACCA---AGATCGAAATTTGTCACAA	237					
Qy	782	aaccagccaagagcgtc-----aggcactgcgaagcggaagcgtgtagtcgacgcatgtc	835					
Db	238	AACAAGATAAACACACGCTGAAAGGCGATGCCCTAGCGACAGGTGAACATCCATGCT	297					
Qy	836	attggtgaaatgctaacgcaagggcggtgtaagccatcgccatcggttagtagtaataa	895					
Db	298	TTTGGTAGTCTTTCTAAGGCACAAAGGCTCTCAAGCTATTGCTATCGTAGTGTCAAAACA	357					
Qy	896	actgtcaatggaagcagcttggataagataggtaaccgatgctacgggtcaagagtcctatc	955					
Db	358	GATCCTAATAATGTGTAGTAATGTAAATGTAGTTCCACGCCAAAGGTAAACAGATCCATC	417					
Qy	956	gccatcggtggtgatgttaaaggctagtggtgatcctcgtattgccatcggttagtgatgac	1015					
Db	418	GCCATCGGTGTGTATTTGGCTGAGGGTGATGCCCTCGATTGCCATCGGTAGTGTATGAC	477					
Qy	1016	ttacattgttgtatcagcatggtaatcctaaacataccgaagggtactctgtattaacgat	1075					
Db	478	TTATATTGGCTAAGAAATCTTGATCTCAAGAA-----TGAATTTACAAA	522					
Qy	1076	cttattaacggcatcagctattaaagaataacgaagctcaagagataatgatgtataaa	1135					
Db	523	CTTATTCATGCCCATGAAATATTAAAAAAAATACAAACCTCAACCCATGGTAAAAATCAA	582					
Qy	1136	tatagcgcacaacccgcaagcgacacgacgagcagctactgcagtgtaggcacgtgcatatga	1195					
Db	583	TATCGACGCACAAGAGCACAAAGGCAGCCAGTACTTGTGAGTGGAGGCCATGTCATATGCA	642					
Qy	1196	cagggtcatttttccaacgcctttgttatacagggcgcaacagctaaaaagtgccctattcctg	1255					











Qy	6311	gaacgcgcgtatccaaaatgtgcagcaggtgaggtcagtgccaccagcaccgattgcgcgttc	6317			
Db	2761	GAACGCCGTATCCAAAATGTGGCAGCAGGTGAGTGTCAGTGCACCCAGCACCGATTCGGTC	2820			
Qy	6371	aatgtagtcagctgtacaaagccaccaaaagcattgccaaagcaaccaatgaacttgac	6430			
Db	2821	AATGGTAGCCAGTTGTACAAAGCCACCCAAAGCATTGCCAAGCAACCAATGAGCTTGAC	2880			
Qy	6431	catgctatccaccacaaacgaaaaataaggccaatgcagggatttcatacagcagatgcgatg	6490			
Db	2881	CATCGTATCACCAAAAGCAATAAGGCCAATGCAGGGATTTCATCAGCATGCAGTGCAGTG	2940			
Qy	6491	gcgtccalgcacaagcctacatctctgcgacatccatggttacccgggggtattgcacc	6550			
Db	2941	GCGTCCATGCACAAGCCTACATTCCTGGCAGATCATCGTGTACCGGGGTATTTCGCACC	3000			
Qy	6551	cacaacgctcaagtgctgcgtggcagtggaactgcgaagcctgcgataaatggtcaatgg	6610			
Db	3001	CACAAACGGTCAAGTGGCGGTGGCAGTGGGACATGTCGAAGCTGTCGGATTAATGGTCAATGG	3060			
Qy	6611	gtatttaaaatcaatggttcagccgcgatacccaaggccatgtagggcggcagtggttgca	6670			
Db	3061	GTAATTAATCAATGTTTCAGCCGATACCCAAAGGCCATGTAGGGCGGCAGTTGGTGCA	3120			
Qy	6671	ggtttccacttttaa	6685			
Db	3121	GGTTTTCACTTTTAA	3135			
RESULT 10						
AX079922						
LOCUS	AX079922	6942 bp	DNA			
DEFINITION	Sequence 10 from Patent WO0107619.					
ACCESSION	AX079922					
VERSION	AX079922.1 GI:13159441					
KEYWORDS	.					
SOURCE	Moraxella catarrhalis.					
ORGANISM	Moraxella catarrhalis Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.					
REFERENCE	1 (bases 1 to 6942)					
AUTHORS	Loomore,S.M.; Sasaki,K.; Yang,Y.P. and Klein,M.H.					
TITLE	Recombinant high molecular weight major outer membrane protein moraxella					
JOURNAL	Patent: WO 0107619-A 10 01-FEB-2001; CONNAUGHT LABORATORIES LIMITED (CA)					
FEATURES	Location/Qualifiers i. 6942 /organism="Moraxella catarrhalis" /db_xref="taxon:480"					
BASE COUNT	2286 a	1621 c	1564 g			
ORIGIN	1471 t					
Query Match						
Best Local Similarity 37.1%; Score 2590; DB 10; Length 6942;						
Matches 4802; Conservative 0; Mismatches 1105; Indels 1269; Gaps						
Qy	542	atgaatcacatctataaagtcattctttaacaaagccacagcacatttatgcgcagtgcca	601			
Db	1	ATGAATCACATCTATAAAGTCATCTTTAAAGCCACAGGCACATTTATGGCGTGCCA	60			
Qy	602	gaglacgcctaaatccacacagcacgaggggggtagctgtgtacagggcaagttggcagt	661			
Db	61	CAGTGGCCAAATCCCAACGCGGAGGAGTAGCAGTAGTAGTACCGCAGGACAGGTGGCAGC	120			
Qy	662	gtatgcactctgagctttgcccgtattggcgcgctgcgtctcgtctcgtatgcagtgcaacg	721			
Db	121	TCTCCTGTCAATCCCGCTGACTCGTGTGGCAGCTCGCTATCCTAGTATCGGTGGCAGC	180			
Qy	722	ctcagtggcgcgtgcttatgctcaaaaaaagatacacaacatcatcgcaattggtgaacaa	781			
Db	181	CTCAATGGCAGTGTCTATGCTCAAA-----ATAATAGCAAGATCGCATTTGGT-----	228			



Db 2194 AGCATTTACCAATGAGTGGCGTACGACACACAGAAAGAGGCAATACAAATCACACAGCGAC 2253  
 Qy 2990 aaagcaaaatccaacgctgcagcattaaatgatatataataacagcgtttaacactaaaa 3049  
 Db 2254 GAAGACAANAATCCAAGCGCGCAGTATCGGTGATATATTAATACAGCGCTTTAACTTAAA 2313  
 Qy 3050 aataaacaacccattgactttgttccacttatgacatttggactttgccaatggc 3109  
 Db 2314 AATAATAGCAACTCCGTTGGCTTGTCTCCACTTATTAACACTGTGTGACTTATCGATGGC 2373  
 Qy 3110 aatgcaaccacccagcaagtaaccatgatataccgctataacaaaacccagtaagtggtatat 3169  
 Db 2374 AATGCCACCCCGTAAAGTAACTTACGATGAACACCAATCAACACAGTAAAGTAACTTAT 2433  
 Qy 3170 gatgtaattggtatgatacaaacattcatctaacagcgcactgatgacaa---taaaaaa 3226  
 Db 2434 GATCTCAATGTGGATGAGAAAACCATTTGAACACAGCGCATATGCGCAACACAAACAA 2493  
 Qy 3227 ctggcgctcaaaaccccaactgaacaaacaaagtctaatggtatatacagcaactaac 3286  
 Db 2494 ATTGGCGTCAAAACACACACACTGACCAACAACAAATGCTAATGGTAAAGCAAC----- 2546  
 Qy 3287 ttaaatgttaactctagtatgaagatgcccctgtttaacgccaagaacatgcgcgaaaaat 3346  
 Db 2547 --CAACTTTAGTACCCCGGATAAGCATGCCCCCTTTAACGCGCAAGAGCATCGCCGAAAAAT 2604  
 Qy 3347 ctaaacaccttagccaaggaattcacaccacaaagggcacagcagacacccgcctacaa 3406  
 Db 2605 CTAACACCCCTAGCAGGAAATTCACACCACCAAGGACAGCAGACACCGCCCTACAA 2664  
 Qy 3407 accttaccgttaaaaggtagatgaataataatgctgatgacgcgcaacgcctaccc 3466  
 Db 2665 ACCTTTAAAGTCAAAAGACGGTGCA-----ACTGATGACGAAACCATCACCGTGGGT 2718  
 Qy 3467 gtgggtcaaaagaacgaataatacaagtcaaacacccttaacactcaaggttgaaaaaggt 3526  
 Db 2719 AAGATGGTACACAAAACGGCAGACCGCTCAACACTCTAAACACTCAAGGTGAAAAACGGT 2778  
 Qy 3527 cttaatataaaacccgacaaaaatggtaacgtttacctttggcattaacaccacacagcgt 3586  
 Db 2779 CTACCGGTGCTACCAATAAAGATGCTACGGTTACCTTTGGCATTAACACCCAAGCGGT 2838  
 Qy 3587 cttaaacccggcaaaagcacc---ctaaacgagcgttgcttctattaaacacccacat 3643  
 Db 2839 CTTAAAGCGCGCAGACACCACTCTAAACAAGATGGCTTGCTATTAAACACCCCGCT 2898  
 Qy 3644 gtagcgaacaaatccaagtcggtgctgagtggtgagtttgccaaaggttaataataat 3703  
 Db 2899 AGTAACGAACAAATCCCAAGTCGGTGGCTGATGGCGTGAAGTTTGCCAAAGGT---TGATAAG 2955  
 Qy 3704 ggtgtgtagtgctggtcattggtgacacactgcgcattaccagagatgaaattggcttt 3763  
 Db 2956 GGTAAATCAAGCACTGGCATTTGATGGCACAAGCCGTATCACCACCAAGATCAAAATGGCTTT 3015  
 Qy 3764 actgggactaatggctcacttgataaagaacaaacccacacttaagcaagac----- 3814  
 Db 3016 ACTGGGCTAATGGCTCAGTGTATACCAACCAACCAACCCACCTTAACCAAGACAAAGCTTAAA 3075  
 Qy 3815 -----ggcattaacgcaggtggttaaaagattaccaac 3847  
 Db 3076 GTGGGTGAAGTTGAATTAACCAACTGGCATTAACCGAGGTGGTAAAGAAATTTACCAAC 3135  
 Qy 3848 attcaatcaggtgagattgcccaaaacagcgcattgctgtgacagcgcgcaagatttat 3907  
 Db 3136 ATTCAATCAGGTGATATTACCCAAACACGCAATGATGCTGTGACAGCGGTGCGGTTTAT 3195  
 Qy 3908 gatttaaaacccgacttgaaacaaatcagcagtagtactgcacaaacagcacaacactca 3967  
 Db 3196 GATTTAAAAACCGAATTGAAAGCAAAATCAACAGTGTGCTTAAACAGCAGCAAAACTCA 3255  
 Qy 3968 ttacacgaattctcagtagcagatgaacaagggttaactttacggttagtaaaccttac 4027

Db 3256 TTACACGAATTCCTCAGTAGCAGATGAACAAAGGTAATCACTTTACGGTTAGTAACCCCTTAC 3315  
 Qy 4028 tccaatttatgacacctcaaaagacctctgatctcatcacttgcaggttgaaaaagcgatt 4087  
 Db 3316 TCCAGTTATGACACTCAAGACCTCTGATGTATCATACCTTTGCAAGTGAANAAGCGCAT 3375  
 Qy 4088 accaccaaggttaataaaggtgtgtgtggcatttgaccaaaacaaagggttaacc 4147  
 Db 3376 ACCACCAAGGTAAATAAAGGTGTGGTGGCATTTGACCAACCAAAAGGCTTAACC 3435  
 Qy 4148 acgcctaagctgacccgtgggttaataataatggcaaaagcattgtcattgacagcaaaaaat 4207  
 Db 3436 ACGCTTAAGCTGACCGTGGGTAAATAAATGCAAGGCAATGTTCATTGACAGTAAAGAT 3495  
 Qy 4208 ggtcaaaatcaccatcacaggaactaagaacactctagctaatgttaccatgataaaggt 4267  
 Db 3496 GGTCANAATACCATCAGAGACTAAGCAACACTCTAGCTATGTATTACCAATGA-----T 3549  
 Qy 4268 agctacgcacccacagacaggggcaataataatcaaaagcagaacaaacccgtgcgcgc 4327  
 Db 3550 GGTGCAGACACGCACTAAAGCAAGGGCTTGCCAAATGACACCGACAAACCGGTGCCGCC 3609  
 Qy 4328 agcattgttaattgcttaagcagcagcgttttaacttgaaggcaatggtgaagcgttgac 4387  
 Db 3610 AGCATTTGGTGTGCTGCTAAACGCGAGGCTTTAACTTTGCAAGGCAATGGTGAAGCGTTGAC 3669  
 Qy 4388 ttgtctccacttatgacacccgtcaactttgctgagtgcaatgcccaccgcgtlaaggtg 4447  
 Db 3670 TTTGCTCCACTTATGACACTGTTGACTTTTATCGATGGCAATGCCACCACCGCTTAAGGTG 3729  
 Qy 4448 acctatgatgacacaaacccagtaaaagtgtgtctatgatgtcaattgtgtgataaca 4507  
 Db 3730 ACCTATGATGACAAAGCAAAACCCAGTAAAGTGGTCTATGATGTCAATGTGGATAATAA 3789  
 Qy 4508 accaattgaagttaa---gataaaaaacttggcgtataaaaaacccacacattgaccagttact 4564  
 Db 3790 ACCAATTGAAGTGACAAGTGATAAAAAAATTGGCGTCAAAACCCACCACTGACCAAAAAC 3849  
 Qy 4565 ggcaaggtgcttaataattggcctgaagcaatcaagctactggtgcgttgctgttcaag 4624  
 Db 3950 AGTCTTAATGTTATGCA---ACCAAAATTTAGTCCCGCGGATGGCGATGCTTGTTTAAA 3906  
 Qy 4625 gccagtgatattcgttgcctatctaacaaccttattgtggcgacatccaactgccaaggg 4684  
 Db 3907 GCCAGTGATATCGCACCCATCTAAATACCTTGGTGGCGACATCCAAACCGCCAAAGGG 3966  
 Qy 4685 gcaagcaaggaacaaactcagcaggtctatgtgagctgctgagcaataaaggtctatctat 4744  
 Db 3967 GCAAGCCAAGCAAGCAGCTCAGCAAGCTTATGTGATGCTGATGGCAACAAGGTTCATCTAT 4026  
 Qy 4745 gacagtcaccgatacaagctactatcaagcgaataatgatggcagctgataaaccgcaaa 4804  
 Db 4027 GACAGTACCGATAGAAGTACTATCAAGTCAATGACAGGGTCAAGTGGACACANAACANA 4086  
 Qy 4805 gaagtggccaaagacaaactggtgcgcaagcccaaacccagatggcacatttggctcaa 4864  
 Db 4087 GAAGTTGCCAAAGACAAACTGGTGGCGCCCAAGCCCAACCCAGATGGCAGCATTTGGCTCAA 4146  
 Qy 4865 atgaatgtcaaatcagtcattacaagaacaaagtaaatgatgccaataaagaacgaagc 4924  
 Db 4147 ATGAATGTCAATCAGTCATTAACAAAGAGCAAGTAAATGATGCCAATTAANAAGCAAGGC 4206  
 Qy 4925 atcaatgaagacacgccttggtaaaggacttgaaaaagcgccttctgataaacaacaaacc 4984  
 Db 4207 ATCAATGAGACACACGCTTTATCAAGGGCTTTGANAACGCGCCCAAGACACCAANAACC 4266  
 Qy 4985 aaaaacccgcagtaactgtgtgggtgatttaaatgcccgttggccaaacaccgcctgaccttt 5044  
 Db 4267 AAAAAACCGCAGTAACCTGTGGGTGATTTAAATGCCGTTGCCCAACACCGCTGACCTTT 4326  
 Qy 5045 gcaggggatacaggaacacacgccttaaaaaactggcgagagactttgacatcaaaaggtgg 5104  
 Db 4327 GCAGGGGATACAGGCACAAACGGCTAAAAAAGTGGCGAGACTTTTGACCATCAAAAGTGGG 4386

QY	5105	caaacagacacccaataagctaaccgataataacatcgggtggtgtagcaggtactgtagc	5164
Db	4387	CAAAACAGACACCAATAAGCTAACCCGATAATAACATCGGTGTGTAGCAGGACTATGCGC	4446
QY	5165	ttcactgtcaaatctgccaagacctaaccatcttaacagcggttaatgcaggtgagcacc	5224
Db	4447	TTTCACTGTCAAACTTGCCTAAAGACCTTAACCAATCTTAACAGCGTTTAATGCAGGTGGCACC	4506
QY	5225	aaaattgatgacaaaggcggtgtcttttttagactcaagcggtcaagccaaagcaaacacc	5284
Db	4507	AGAAATGTATGAAAAAGGCACTCTCTTTTGTAGACGCAAAACGGTCAAGCCCAAAACACACC	4566
QY	5285	cctgtactaaatgccaatggcgtgactgactgggtggc-----	5320
Db	4567	CCTGTGCTAAGTGCCANTGGCTGGACCTGGGTGGCANNCGCATCAGTAACATCCGTGCA	4626
QY	5321	-----	5320
Db	4627	GCTGTGTGATGATAACGATCGGTGAACCTTTAAGCAGTTTAATGAAGTTGCCAAAACGGTCC	4686
QY	5321	-----	5320
Db	4687	AACAACCTAAACAACCAAAAGTAACTCAGTGGCTCATTTACCCTTTGTGTAAACCGATGCC	4746
QY	5321	-----	5320
Db	4747	AATGGCAAGCCCATCAATGCCACCGATGCAAGCCCAAAAGCCCATCAAGGGCGCGAT	4806
QY	5321	-----	5320
Db	4807	GGTAAATACTATCACGCCAACGCCAACGGGTACCTGTGGCAAAAGATGGCAAGCCCATC	4866
QY	5321	-----	5320
Db	4867	ACCGATGGCGGACAAACTTGGCAATCTGGCAGCTCATGTGGCAAAACCCCTTGATGCAGGTCA	4926
QY	5321	-----	5320
Db	4927	CAAGTGTGGCAAGCCTAGCGCGCAACTCAGATGCCATCACCTCAACCAACATCAAGTCC	4986
QY	5321	-----	5320
Db	4987	ACTTTGCCACAAATTGACACACCAACACAGAGTAATGCCAATGCCAGGGCAAGCCCAAGT	5046
QY	5321	-----	5320
Db	5047	CTGCCAGCCTATCAGCAGCACAGCAAAAGTAATGCTGCCAGTGTCAAAGATGTGCTAAAT	5106
QY	5321	-----	5320
Db	5107	GTAGGCTTTAACTTGACAGACCAATCAAAATCAAGTGGACTTTGTCAAAGCCTATGATACC	5166
QY	5321	-----	5320
Db	5167	GTCAACTTTGTCAATGTGTACAGGTGGCGACATCACAGCGTGGTAGTGTGTATGCCACG	5226
QY	5321	-----	5320
Db	5227	ATGAGTAACATCACCGTCAACACCGCCTTAGCAGCGACCGATGATGATGGCAATGTGCTT	5286
QY	5321	-----	5320
Db	5287	ATCAAGCCAAAGATGGTAAGTTCTACAAAGCAGACACCTCATGCCAAGCGCTCACTA	5346
QY	5321	-----	5320
Db	5347	AAAGCAGCAAAATCAGCCAGTGTATGCCAAAACCTCAACTGGTCTTAAGCCCTTGTTAACCCC	5406
QY	5321	-----	5320
Db	5407	AATGCTGGTAAAGCAGTACAGCGGATGCAGTGGCTCTTAATAAATCTTATCAAAAGCGGTA	5466
QY	5321	-----	5320
Db	5467	TTTAAATCCAAAGATGTACAACTACTTACCACAGTAAGCTCTGATGCATCAGTATCCAA	5526
QY	5321	-----	5326
Db	5527	GGCAAAGATAACACAGCATCACCTTAAGCAAAGATGGGTGAATGTAGGGCGTAAAGTTC	5586
QY	5327	atcagtaattgtgggcaaaaggcacaacaaagataccgacgctgccaatgtacaacagcttaaac	5386
Db	5597	ATCAGCATGTGGGTAAAGGCACAAAGACACCGACGCTGCCAATGTACAAACAGTTAAAC	5646
QY	5387	gaagtacgcgaactgtgtgggtcttgggtaattgtgttaataatgataacgctgaacggaatacag	5446
Db	5647	GAAGTACGCAAACTTGTGGGTCTTTGGTAATGCTGTATGATAACGCTGACGGCAATCAG	5706
QY	5447	gtaaacattgcccgcacatacaaaaagaccaaattcagattcatcatctaaacgcactgtc	5506
Db	5707	GTAACATATGCCGACATCAAAAAGACCCAAATTCAGGTTTCATCTTAACCGCAGCTGTC	5766
QY	5507	atcaaacgagcacggttacttggcggttaaggtaataacgataccgaaaaaacttgcoact	5566
Db	5767	ATCAAAAGCAGCACGGTACTTTGGCGGTAAAGTAATAACGATACCGMAAAACTTGGCACT	5826
QY	5567	ggtggtatacaagtggtgggtggataaaagacggaacgcaacgctaacgagatttaagcaatgtt	5626
Db	5827	GGTGTGTACAAAGTGGCGGTGGATAAAGACGCAACGCTAACGGCGATTTAAGCAATGTT	5886
QY	5627	tggtgcaaaaccccaaaaagataggcagcaaaaagccctgctgcacactataaacgccca	5686
Db	5887	TGGGTCAAAACCCAAAGATGGCAGCAAAAGCCCTGCTCGCCACTTATAACGCCCA	5946
QY	5687	ggtcagacaaactattgacaaacaccccgaggaagccattgacagataaaatgaacaa	5746
Db	5947	GGTCAGAGCAACTATTTGACCAACACCCCGCAGAAAGCATTTGACAGAAATAAATGAACAA	6006
QY	5747	ggtatccgtctcttcactgtcaacgatggcaatcaagagcctgtgttcaagggcgtaac	5806
Db	6007	GGTATCCGCTTTTCCATGTCAACGNTGGCAATCAAGAGCCTGTGTACNAGGGCGTAAC	6066
QY	5807	ggcattgactcaagtgcctcagggcagcactcagtggtggataggtttccagggccaaagca	5866
Db	6067	GGCAITGTACTAAGTGCCTTCAGGCAAGCACTCAGTGGCGATAGGTTTCCAGGCCAAGGCA	6126
QY	5867	gatgtgaaagccgctgtgccaatagcagacaaaccccaagcagcaacaaatccatccgc	5926
Db	6127	GATGTGAAGCCGCGTGGCCATAGCGACAAACCCCAAGCAACCCCAAGCAACCAATCCATCGCC	6186
QY	5927	atcgggtgaacgcgcaacgacgagcgatcaatccatcgccatcggttacagggcaatgtg	5986
Db	6187	ATCGGTGATAACGACAAAGCCAGCGGATCAATCCATCGCCATCGGTACAGGCAATGTG	6246
QY	5987	gtagcaggtaagcactctgtgtccatcgcgacccaagcactgttaaggctataacagt	6046
Db	6247	GTAACAGGTAAAGCACTCTGTGCCATCGCGCACCCCAAGCACTGTAAAGGCTCATACAGT	6306
QY	6047	tacagtgtgggttaataaacacccagtttaccgatgccactcaaacccgatgtctttgggtg	6106
Db	6307	TACAGTGTGGGTAAATAACCAACCAAGTTTATCGATGCCACTCAGACCGATGCTCTTGGTGTG	6366
QY	6107	ggcaataacacatcacccgtgaccggaagtaactcgggttgccttaggtttcaaacctctgcac	6166
Db	6367	GGCAATAACATCACCTGACCGAAAGTAACCTCGGTTCGCTTAGGTTCAAACTCTGCGCATC	6426
QY	6167	agtgcaagcacacacgagcac	6226
Db	6427	AGTGCAGGACACACACGACGACACACACACACACACACACACACACACACACACACAC	6486
QY	6227	accacagcaggtgcaacccggtacggttaagggcttgcgtggacaaacggttgggtg	6286
Db	6487	ACCACAGCAGGTGCAACAGGTACGGTTAAAGGCTTTGCTGGACAAACGGCGGTGGTGCG	6546
QY	6287	gtctccgtgggtcctcaggtgctgaacgccgtatccaaaatgtggcagcaggtgagggtc	6346

Db	6547	GTCTCCGTGGTGCTCAGGTGCTGAACGCGGTATCCAAATGTGGCAGCAGGTGAGGTC	6606
Qy	6347	agtgccaccagcaccgcatgcggtgctcaatggttagcaggtgtgtaaaagccacccaaagcatt	6406
Db	6607	AGTGCCACACGACCGATGGGTCAATGGTAGCCAGTGTACAAAGCCACCCAGGCATT	6666
Qy	6407	gccaacgcaaccaatgagcttgaccatcgatccaccaacaaacgaaaaataaggccaatgca	6466
Db	6667	GCCAAACCAACCAATGAGCTTGACCATCGTATCCACCAAAACGAAATATAAGCCAAATGCA	6726
Qy	6467	gggaatttcacagcatgagcgatggcgctccatgcccaagccctacattcctggcagatcc	6526
Db	6727	GGGATTTTCACAGGATGGCGATGGGCTCCATGCCAAGCCCTACATTTCTTGGCAGATCC	6786
Qy	6527	atggttaccgggggtattgcccaccacaacggtcaagtgctggtggcagtggtgactgctg	6586
Db	6787	ATGTTTACC GGGGTATTGCCACCCACACGCTCAAGTGGCGGTGGCAGTGGGACTGTGC	6846
Qy	6587	aagctgtcgataatggtcaatgggtatttaaaatcaatggttcagccgatacccaagcc	6646
Db	6847	AAGCTGTGCGAATAATGGTCAATGGGTATTTAAAAATCAATGGTTTCAGCCGATACCCAAGGC	6906
Qy	6647	catgtaaggcgcgagcttggtgaggttttcacttt	6682
Db	6907	CATGTAGGGCGGAGTGTGTGAGGTTTTTACATTT	6942
RESULT 11			
AX079959			
LOCUS	AX079959	2448 bp	DNA PAT 22-FEB-2001
DEFINITION	Sequence 47 from Patent WO0107619.		
ACCESSION	AX079959		
VERSION	AX079959.1 GI:13159466		
KEYWORDS			
SOURCE	Moraxella catarrhalis.		
ORGANISM	Moraxella catarrhalis		
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.		
AUTHORS	1. (bases 1 to 2448)		
TITLE	Loosmore, S.M., Sasaki, K., Yang, Y.P. and Klein, M.H.		
JOURNAL	Recombinant high molecular weight major outer membrane protein of moraxella		
FEATURES	Patent: WO 0107619-A 47 01-FEB-2001;		
source	CONNAUGHT LABORATORIES LIMITED (CA)		
BASE COUNT	Location/Qualifiers		
ORIGIN	1..2448		
	/organism="Moraxella catarrhalis"		
	/db_xref="taxon:480"		
	826 a	569 c	556 g 497 t
Query Match 35.1%; Score 2448; DB 10; Length 2448;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	3551	ggtaacgttaccttggcattacacacacagcggtcttaagccggcgaagcaccta	3610
Db	1	GGTACGGTTACCTTGGCATTATACACCAAGCGGCTTTAAAGCGGCAAAAGCACCTTA	60
Qy	3611	aacgacggttgcttgcattataaaacccactggtgtagcgaacaaatccaaagtcggtgc	3670
Db	61	AACGACGGTGGCTTGTCTATTAAACCCACCTGGTAGCGAACAATAATCCAAAGTCGGTGCT	120
Qy	3671	gatggcgtaagttggcaagggttaataatggtgtgttagtgctggtcattgatggc	3730
Db	121	GATGGCGTGAAGTGTGCCAAGGTTAATAATGTTGTTAGTGTGCTGGCATTGATGGC	180
Qy	3731	acaactgcattaccagagatgaattggctttactgggactaatggtcacttgataaa	3790
Db	181	ACAACTGCATTACCAGAGATGAATTTGGCTTTACTGGGACTAATGGCTCATTGTATATA	240

Qy	3791	agaaaaccccccaataagcaaaagacggcattaaacgaggttgtaaaaaagattaccaacatt	3850
Db	241	AGCAAAACCCACCTTAAGCAAAAGACGGCATTAACGACGTTGTTAAAAAGATTACCAACATT	300
Qy	3851	caatcaggtgagattgcccacaaacagcatgactgctgacagcgcaagatttatgat	3910
Db	301	CAATCAGGTGAGATTGCCAAAACAGCCATGCTGTGACAGCGGCAAGATTATATGAT	360
Qy	3911	ttaaaaaccgaaacttgaaaaacaaatcagcagctactgccaacaaacagcacaaaaactcta	3970
Db	361	TTAAAAACCGAATTTGAAACAAAAATCAGCAGTACTGCCAAAAACAGCACAAATCATTA	420
Qy	3971	caggaattctcagtagcagatgaacaaaggttaataactttacggttagtaacccctactcc	4030
Db	421	CACCAATTTCTCAGTAGCAGATGAACAAGGTAAATAACTTTTACGGTTAGTAACCCCTTACTCC	480
Qy	4031	agttatgacacctcaagacctctgctgcatcacctttgcagdtgaaacggcgatcacc	4090
Db	481	AGTTATGACACCTCAAGACCTCTGATGTCATCACCTTTGTCAGGTGAANAACGGCATTACC	540
Qy	4091	accaaggtaaataaagtggtgctgctggtgacattgacccaacaaaggtttaaaccacg	4150
Db	541	ACCAAGGTAAATAAAGGTGTGGTGGCATTTGACCAACCAACCAAGGCTTAACCAACG	600
Qy	4151	cctaagctgacctgggtgataataatggaagagcattgtcattgacagcccaaaatggt	4210
Db	601	CCTAAGCTGACCGTGGGTAATAATAATGGAAGGCATTTGTCAATTGACAGCAAAATGCT	660
Qy	4211	caaaataccatcagcagtaagaacaacactcagctaatgttaccaatgataaaggtgac	4270
Db	661	CAAAATACCATCAGGACTAAGCAACACTCTAGCTAATGTGTACCAATGATAAAGGTAGC	720
Qy	4271	gtacgcaccacagaacaggcgcaataataacaaagcaagaacaaacccgtgcccgcagc	4330
Db	721	GTACGCACACAGAACAGGCGCAATATAATCAAGACCAAGACAAACCCGTGCCGCCAGC	780
Qy	4331	attgttgatgctgaagcaggcgttttaacttgcaaggaatggtgaagcggttgacttt	4390
Db	781	ATTGTTGATGTCTAAGCGCAGGCTTTAACTTTGCAAGGCAATGTTGAAGCGTTGACTTT	840
Qy	4391	gtctccacttatgacacggtcaacttgcgagtgagcaatgccaccacgcgttaagtgacc	4450
Db	841	GTCTCCACTTATGACACCGTCAACTTTGCCGATGGCAATGCCACCGCTTAAGGTGACC	900
Qy	4451	tatgatgacacaaagcaaaacagtaaaagtgctctatgatgtcaatggtgatacacaac	4510
Db	901	TATGATGACACAAACCAACCAAGTAAAGTGTCTATGATGTCATGTGGATGATACAAAC	960
Qy	4511	attgaagttaaagataaaaaacttggcgtaaaaaaccaccacattgaccagttactggcaca	4570
Db	961	ATTGAAGTTAAAGATAAAAAAATTTGGCGTAAAAAACCCACCATTTGACCACTACGACACA	1020
Qy	4571	ggtgctaataaatttgcctaaagcaatcaagctactcagctgctgctgtaagggcaggt	4630
Db	1021	GGTGTAAATAAATTTGCCCTAAGCAATCAAGCTACTGGCGATGGCTTGTCAAGGCCAGT	1080
Qy	4631	gatactgtctctatcaaacaccttatctggcgacatccaaactgccaagggggcgaagc	4690
Db	1081	GATATCGTTGCTCATCTAAACACCTTATCTGGCGACATCCAAACTGCCAAAGGGGCAAGC	1140
Qy	4691	caagcgaaactcagcaggctatgtgtagtctgtaggcaataaaggttcactatgacagt	4750
Db	1141	CAAGCGAAACAATCAGCAGGCTATGTGATGCTGATGCAATAAGGTCATCTATGACAGT	1200
Qy	4751	acgatacaagttactatcaagccaaaaatgtaggcacagttgtataaaccacaaagaagtt	4810
Db	1201	ACCGATAACAAGTACTATCAAGCCAAAAATGATGGCAGTTGTATAAACCCAAAGAAGTT	1260
Qy	4811	gccaagacaaaactgctgcaccaagcccaaacccagatggcacattggctcacaatgaat	4870
Db	1261	GCCAAACACAAACTGGTGGCCCAAGCCCAACCCAGATGGCACAATTTGGTCAATAAT	1320
Qy	4871	gtcaaatcagtcattatacaaaagaaacaaagtaaatgatgccataaaaaagcaaggcatcaat	4930





```
RESULT 13
AE006194/c
LOCUS
DEFINITION
  AE006194 10533 bp DNA BCT 08-MAR-2001
  Pasteurella multocida PM70 section 161 of 204 of the complete
  genome.
ACCESSION
  AE006194 AE004439
VERSION
  AE006194.1 GI:12721958
KEYWORDS
  SOURCE
  ORGANISM
    Pasteurella multocida.
    Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
    Pasteurella.
    1 (bases 1 to 10533)
    May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and
    Kapur,V.
    Complete genomic sequence of Pasteurella multocida, PM70
    21145866
    2 (bases 1 to 10533)
    Zhang,Q. and Kapur,V.
    Direct Submission
    Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
    University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
    55108, USA
FEATURES
  source
    1..10533
    /organism="Pasteurella multocida"
    /strain="PM70"
    /db_xref="taxon:747"
    91..804
    /gene="PM1566"
    91..804
    /gene="PM1566"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AAK03650.1"
    /db_xref="GI:12721959"
    /translation="MTKLHLWIFLALPINALNPOVDNKLQHNQVQSAEVKFEIKDKD
    ELQVTLVLOEGKDALNQLIVARNQALDVAKKHVTQVVSQOQRYTOVRYGKEGQ
    NGWIDRAEVLVQSODPMLSKLVLAEHNFHFKIGGMVATVSEESLARVEKEMTFAVLQK
    FEQKALQSLMKAGKRYLIELDLSPHDVGSTRSVAVSMKSKDAFYSSADHEMALE
    GKTKLRVAVQAELIND"
    868..1467
    /gene="fkpA"
    /note="PM1567"
    868..1467
    /gene="fkpA"
    /codon_start=1
    /transl_table=11
    /product="fkpA"
    /protein_id="AAK03651.1"
    /db_xref="GI:12721960"
    /translation="MKIAKNVVSIAQVYRTEDGVLVDEAPVNPQLEYLQGHNNLVIG
    LNALGKAVGDFEVRPEEAYGEINENMVORPKDVFQGVDELVLVGMRFITADYDI
    GPUPVITVEAGNVDVVDNMLAGQELLFSVSVATREATLEEIAHGHIEGGGCGG
    GHHDSDREGCHGCGSHHHEHHEHHAHDGCGGCGCKH"
    complement(1521..1733)
    /gene="PM1568"
    complement(1521..1733)
    /gene="PM1568"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AAK03652.1"
    /db_xref="GI:12721961"
    /translation="MPEFRDLISKLEDAHNRLFEKHNELDHRIKNIEANIELGT
    POEVNKKKEKLKDEIYVILKKH"
    complement(1830..2300)
    /gene="yfkJ"
    /note="PM1569"
    complement(1830..2300)
    /gene="yfkJ"
    /codon_start=1
    /transl_table=11
    /product="yfkJ"
    /protein_id="AAK03653.1"
    /db_xref="GI:12721962"
    /translation="MTIHLFVCLGNICRSPMAEFLMRKIKOAGLETIVTKTSAGTS
    GMDHGDHMGCTADILDQCCHIDSTGFRSKRVQHQWQEDYIIAMDNSNLDLRLFG
    HHPDKLFKITELCPDLGHPIDPWYSKDFNPTQYLLNPPCCDALLARLEKHL"
    complement(2415..6314)
    /gene="hsf_2"
    /note="PM1570"
    complement(2415..6314)
    /gene="hsf_2"
    /codon_start=1
    /transl_table=11
    /product="Hsf"
    /protein_id="AAK03654.1"
    /db_xref="GI:12721963"
    /translation="MNIYRTLNAAATQSWVVVSELAKAGKKSASKSALVNSVSGFS
    FTLIAASVVLGSGQVNAAEVTGNTGVSDDKYCFYNASSQSVICGDTATTKTDTONK
    PAKSVVIGCATNDGETNVAIGAKSSKSAASIAIGDNKALDNQAIAGONATANSID
    WDISIGRGAQEOTEVSAEGRNIAIGDGALRGKGVNNIALCTGSAGDRLAGTHNVLM
    GTVNADEAVRSALTAGSTTKEINAKETTDNKYYIEASNTVALGTALATQALAAVIG
    QQAKFGNSVAVNGTKASGQTATGATGSSNAHATGSSIAIGTIVNGDNTTTLASDN
    FTIAMLSAQATKSDAIAIGRNAKAGINTVSGINAGVTCTADEAQSDPNKLEPAT
    DAVFIGNAGYKSNQRMQVSLGDSGEVGTENTVIGNSAGNTKGNNTNVAISSRA
    GONVEGHDNFAALIEAGONTIGSDNIAIGHAGRSADPNTKLNINTTISLCKESVSLK
    NFGIAOONKAKTDLGASIAIGRNAEAVGGETANIAIGDSASADASGAIVLGTQAOKS
    LTVDGKKYGAYSIVIGTEAKIAQAAPAGKKNPKDAIAIGTKEAHYASTALTALFG
    AKSDTKAAYSGISGNKGAIAFGSEAKTTENAGSSIAFGTGAQTRASASTAIGM
    GAETGDFGQALGSDAVALGREAKAKQNALAFYKADVHKDVALGAGAEATAAE
    GTNEATVNEFKYSGFAGIKPIATVSVGKKDAERTITNVAAGRIDKTSTDAINGSQLYL
    ALNALGVNGLTVNLVNGDAATVKGDEAGTLMNSIGTGKGTTHDAIEAVNNTAK
    AKSTTVKEGDNITVTEAADSGRTVTATKDVKDFSVVAGCTKIDANGLTFVDDQG
    TKIDNTPSISKTGIDAGNQKVTNVQNGNTAKDSKDAVNGCOLPAQGEYKNIIGDGT
    YNPATEKVTNTDGTGASTIDEIRAKVNTTAKAEVTVQGENIVVTSAPGANGTV
    YVATATAEFTDGTGVSVDITNDITGLSNKTLGGDNFNAKNGRAASBEQLNATQT
    NLATLVGNAQNTNGNVMTDIDGCTGKNKINDAIKASRNEVKQGNKNNVPTTGANGQ
    TIYEATADKAVDFEVKVGITIDATNTKISGAKGDISENSTDAVNGSOLYELQOKI
    AKGSDNINLNNRINKVDKDLRAGIAGANAAGLPQAYIPGKSMVAVAAGTYKQGNAI
    ALGMSRLSDGKVIILKLTGNTNSRGDEGASIGAGYOW"
    complement(6493..7833)
    /gene="PM1571"
    complement(6493..7833)
    /gene="PM1571"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AAK03655.1"
    /db_xref="GI:12721964"
    /translation="MTSSLPNIGFISLIGCPKNLVDSEIRILTELRSBGYNIPSVENAD
    LVIVNTCGFIDSAVQESLEAIGALEENGKVIIVTGLCAKEDRIREVHPKVLVETGPH
    SYEAVNQVHKYVPKPAYNPVNLVPKQGVKLPKHAYALKISEGCDHRCFTCLIPSM
    RGDLDSRSTQVLDEAKRLVEAGVEKLVVSQDTSAYALDRSKEEQNTVFWNGMPKIK
    NNLISCEQOGLKGLVWVHYVYPHVDQLIPMAEGKILPYLDLIPLOHASPKILKA
    MKPISCIERTLERIKKREICPDLTIRSTIVFGPGSESEDFQLLDLFLKALLDRVG
    CFKFSVEGAIATAMPQDVPVEEKARFHRFMOLQOEISAAALQOKVGKVTVLVDEV
    DEEGIIARSIADAPEIDGVVYIDNPNRVAVKAGQFIEVKITRADAYDLASLIN"
    complement(8029..8787)
    /gene="udp"
    /note="PM1572"
    complement(8029..8787)
    /gene="udp"
    /codon_start=1
    /transl_table=11
    /product="udp"
    /protein_id="AAK03656.1"
    /db_xref="GI:12721965"
    /translation="MSEVPHLGLTKAMLEGATLAIIPGAPERVEKIAKLDLRPKFLAS
    TREFTSLWLVIGDKAVVVCSTGIGGSPVSIATVEALQGLGVRTFLRTGTTCAIQPHINV
    GDILVTTGAVRLDGASLHFAPMEYPAVANFECTNALLYKAATELANOKVYVGTAAADST
    FYPGQERYDTPYSGVYHFGSLKQODLNVMFEMESATLFTMCSALGRAGWVAGA
    IVNRTOOEIPNEAAVKDIEKNAVEIVVKAAYLL"
```







---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 12:39:38 ; Search time 1369.14 Seconds  
(without alignments)  
3197.431 Million cell updates/sec

Title: US-09-361-619-5

Perfect score: 6972

Sequence: 1 ccattgatattggcagggtgt.....ccgttgacctgcagggtgac 6972

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_0601.\*  
1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	6972	100.0	6972	22	AAF59102 M. catarrhalis str
2	6961	99.8	6973	22	AAF59100 M. catarrhalis str
3	6959	99.8	6971	17	AAT38740 Moraxella outer me
4	6258	89.8	6259	22	AAF59106 M. catarrhalis M56
5	6144	88.1	6144	22	AAF59103 M. catarrhalis str
6	5979	85.8	5979	22	AAF59101 M. catarrhalis str
7	4671.2	67.0	62909	22	AAF28545 Genomic fragment #
8	4101.6	58.8	6159	22	AAF59104 M. catarrhalis str
9	2590	37.1	6942	22	AAF59105 M. catarrhalis les
10	720	10.3	720	22	AAF59129 M. catarrhalis str
11	426.8	6.1	1000	22	AAF591426 Moraxella catarrha

12	148	2.1	3030	21	AA92496	Haemophilus influe
13	146.4	2.1	3036	21	AA92493	Haemophilus influe
14	141.6	2.0	3354	21	AA92495	Haemophilus influe
15	131.2	1.9	3342	21	AA92498	Haemophilus influe
16	125.2	1.8	936	22	AAF58252	Oligonucleotide D1
17	125.2	1.8	936	22	AAF58254	Oligonucleotide D1
18	125.2	1.8	936	22	AAF58257	Oligonucleotide D1
19	125.2	1.8	936	22	AAF58259	Oligonucleotide D2
20	125.2	1.8	936	22	AAF58262	Oligonucleotide D2
21	125.2	1.8	938	22	AAF58255	Oligonucleotide D1
22	124.8	1.8	936	22	AAF58252	Oligonucleotide D1
23	124.8	1.8	936	22	AAF58254	Oligonucleotide D1
24	124.8	1.8	936	22	AAF58257	Oligonucleotide D1
25	124.8	1.8	936	22	AAF58259	Oligonucleotide D2
26	124.8	1.8	936	22	AAF58262	Oligonucleotide D2
27	124.8	1.8	938	22	AAF58255	Oligonucleotide D1
28	85.4	1.2	7486	19	AAV22837	Haemophilus para
29	83.8	1.2	2079	21	AA92494	Haemophilus influe
30	83.8	1.2	8930	19	AAV22834	Haemophilus para
31	83.6	1.2	2037	17	AAT41477	Haemophilus adhesi
32	82.8	1.2	3294	17	AAT41475	Haemophilus adhesi
33	78.2	1.1	3300	21	AA92497	Haemophilus influe
34	78.2	1.1	7253	21	AA92499	Haemophilus influe
35	78.2	1.1	7291	17	AAT41476	Haemophilus adhesi
36	76.8	1.1	1830121	17	AAT42063	Haemophilus influe
37	67	1.0	1757	21	AA92500	Haemophilus influe
38	61	0.9	2308	20	AA92500	DNA encoding a sur
39	60.2	0.9	1797	20	AA92500	DNA encoding a sur
40	60	0.9	60	17	AAT38741	Moraxella outer me
41	59.4	0.9	102634	21	AA81464	N. meningitidis pa
42	59.4	0.9	349980	21	AAF21609	Neisseria meningit
43	59.4	0.9	1437668	21	AA81490	N. meningitidis B
44	58.8	0.8	244	22	AAF58238	Oligonucleotide D1
45	58.6	0.8	1797	20	AA92500	DNA encoding a sur

#### ALIGNMENTS

RESULT 1

AAF59102

ID AAF59102 standard; DNA: 6972 BP.

XX AAF59102;

XX AC

XX 24-APR-2001 (first entry)

XX DT

XX DE

XX M. catarrhalis strain 4223 genomic 200kDa gene SEQ ID NO:5.

XX Moraxella catarrhalis strain 4223; major outer membrane protein;

XX 200kDa outer membrane protein; antibacterial; immunogenic; infection;

XX Moraxella catarrhalis.

XX WO200107619-Al.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX P-PSDB; AAB69134.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

XX useful in protective vaccines and for diagnosis

XX



Db 1801 tggggtgaaacacactacataaataagcccaacacaggttaaggtaggttagtagtag 1860  
Qy 1861 tactacagctgaattattgagtgatagtttaacctttaccagcccaataacagcgagtc 1920  
Db 1861 tactacagctgaattattgagtgatagtttaacctttaccagcccaataacagcgagtc 1920  
Qy 1921 aagcacaagcaaaacccgctctatggcgttaattgggtggaagtttacttaataatcaga 1980  
Db 1921 aagcacaagcaaaacccgctctatggcgttaattgggtggaagtttacttaataatcaga 1980  
Qy 1981 aacagcagcaatcgccactactcgattaccagagataaataaggcttgcgcagatgg 2040  
Db 1981 aacagcagcaatcgccactactcgattaccagagataaataaggcttgcgcagatgg 2040  
Qy 2041 tgaattgataaataaacaagcaccatatttgataaataaacaacttaaaagtggttagt 2100  
Db 2041 tgaattgataaataaacaagcaccatatttgataaataaacaacttaaaagtggttagt 2100  
Qy 2101 tgaattaccatgacaaatggcgttgcagtgatgcaggttaataaataagatcaatttgc 2160  
Db 2101 tgaattaccatgacaaatggcgttgcagtgatgcaggttaataaataagatcaatttgc 2160  
Qy 2161 agtagcagtgctaaagcagtggttaccatcgaaacagctcaagccgccaagcctactt 2220  
Db 2161 agtagcagtgctaaagcagtggttaccatcgaaacagctcaagccgccaagcctactt 2220  
Qy 2221 aaacgagcgctggcagtcagtcacacactgaaatacagttgagtcgaagtggtg 2280  
Db 2221 aaacgagcgctggcagtcagtcacacactgaaatacagtcgaagtggtg 2280  
Qy 2281 caattgtaccgcccacacttacacattggcgtgaaacccacgagcttaacagtgatgg 2340  
Db 2281 caattgtaccgcccacacttacacattggcgtgaaacccacgagcttaacagtgatgg 2340  
Qy 2341 cactagtgataaatttagtgtaaggttagtgtagcagaacaaatagcttagttaccgcca 2400  
Db 2341 cactagtgataaatttagtgtaaggttagtgtagcagaacaaatagcttagttaccgcca 2400  
Qy 2401 acatttggcagctatctaataagtcgaatcgaaacgctgacagtgctctacaaagctt 2460  
Db 2401 acatttggcagctatctaataagtcgaatcgaaacgctgacagtgctctacaaagctt 2460  
Qy 2461 taccgttaagaagaagacagatgatgacgccaacgctatcacccgtggttaagaatacag 2520  
Db 2461 taccgttaagaagaagacagatgatgacgccaacgctatcacccgtggttaagaatacag 2520  
Qy 2521 aaaaaatggcgcagtcagcatcttaaaactcaaaaggttaaaaaaggttaaacggttgc 2580  
Db 2521 aaaaaatggcgcagtcagcatcttaaaactcaaaaggttaaaaaaggttaaacggttgc 2580  
Qy 2581 taccaaaaaagatggtagcgttaccttggcgttagccaagatagcgggtgacacattgg 2640  
Db 2581 taccaaaaaagatggtagcgttaccttggcgttagccaagatagcgggtgacacattgg 2640  
Qy 2641 caaaagcaccctaaacacagatggcttgaactgttaagaatacacaagcaaaaatcccaagt 2700  
Db 2641 caaaagcaccctaaacacagatggcttgaactgttaagaatacacaagcaaaaatcccaagt 2700  
Qy 2701 cgggtcgaatggcattaaatttactaattgaaatggtagtgtaataccaggtactggcattgc 2760  
Db 2701 cgggtcgaatggcattaaatttactaattgaaatggtagtgtaataccaggtactggcattgc 2760  
Qy 2761 aaataccgcctcgcatcaccagagataaaattggcttgcgttgcgttctgagtggtgcagttga 2820  
Db 2761 aaataccgcctcgcatcaccagagataaaattggcttgcgttgcgttctgagtggtgcagttga 2820  
Qy 2821 tacaacaaaccccttatcttgataagacaagctacaagttggcaatgttaagattaccaa 2880  
Db 2821 tacaacaaaccccttatcttgataagacaagctacaagttggcaatgttaagattaccaa 2880  
Qy 2881 cactggcattaaagcaggttgtaaaagccatcacagggtgtcccccaacactgcttagcat 2940  
Db 2881 cactggcattaaagcaggttgtaaaagccatcacagggtgtcccccaacactgcttagcat 2940

Db 2881 cactggcattaaagcaggttgtaaaagccatcacagggtgtcccccaacactgcttagcat 2940  
Qy 2941 tgcgcatacaagtagccgcaacatagaaactgggcaatacaatccaagacaagaacaatc 3000  
Db 2941 tgcgcatacaagtagccgcaacatagaaactgggcaatacaatccaagacaagaacaatc 3000  
Qy 3001 caacgcgtccagcagcattaatgatataataatacaggttttaacctaaataataaaca 3060  
Db 3001 caacgcgtccagcagcattaatgatataataatacaggttttaacctaaataataaaca 3060  
Qy 3061 cccattgacttgcctccacttatgacattgttgacttgccttgccttgccttgccttgc 3120  
Db 3061 cccattgacttgcctccacttatgacattgttgacttgccttgccttgccttgccttgc 3120  
Qy 3121 cgcacagtagtaaccatgataccgctaaacaaacagtaaaagtggtatagtgaaatgt 3180  
Db 3121 cgcacagtagtaaccatgataccgctaaacaaacagtaaaagtggtatagtgaaatgt 3180  
Qy 3181 ggaatgatacaacacattctatacagcagcactgatacaataaaaaaacttggcgtcaaaa 3240  
Db 3181 ggaatgatacaacacattctatacagcagcactgatacaataaaaaaacttggcgtcaaaa 3240  
Qy 3241 cacaacactaaacaaacaaagtgctaatggttaatacagcaactaactttaatgtaactc 3300  
Db 3241 cacaacactaaacaaacaaagtgctaatggttaatacagcaactaactttaatgtaactc 3300  
Qy 3301 tagtgatgaagtgcccttggtaacgcgaagacatcgccgaataatctaaacacccctagc 3360  
Db 3301 tagtgatgaagtgcccttggtaacgcgaagacatcgccgaataatctaaacacccctagc 3360  
Qy 3361 caaggaaattcacaccacccaaagggcacagcagcacccgcccatacaaaacttaccgttaa 3420  
Db 3361 caaggaaattcacaccacccaaagggcacagcagcacccgcccatacaaaacttaccgttaa 3420  
Qy 3421 aaaggtagatgaaaaataataatgctgtagcgcacaaagccatcacccgtgggtcaaaagaa 3480  
Db 3421 aaaggtagatgaaaaataataatgctgtagcgcacaaagccatcacccgtgggtcaaaagaa 3480  
Qy 3481 cgcataataacagtcacaccccttaacactcaaaaggtgaaacggttcttaataataaaac 3540  
Db 3481 cgcataataacagtcacaccccttaacactcaaaaggtgaaacggttcttaataataaaac 3540  
Qy 3541 cgcataaaaaatggtagcgtttaccccttggaattacacacacacggttcttaaaagccgcaa 3600  
Db 3541 cgcataaaaaatggtagcgtttaccccttggaattacacacacacggttcttaaaagccgcaa 3600  
Qy 3601 aagcaccctaaacagcagcgttggcttctattaaaaaccccaactggtagcgaacaaatcca 3660  
Db 3601 aagcaccctaaacagcagcgttggcttctattaaaaaccccaactggtagcgaacaaatcca 3660  
Qy 3661 agtcgggtgctgtagcgttgaaggttgcgaaggtttaaataatgggtgtgttaggtgctgg 3720  
Db 3661 agtcgggtgctgtagcgttgaaggttgcgaaggtttaaataatgggtgtgttaggtgctgg 3720  
Qy 3721 cattgtatggcacaactcgcatcaccagatgaaatgggttcttactgggactaagcttc 3780  
Db 3721 cattgtatggcacaactcgcatcaccagatgaaatgggttcttactgggactaagcttc 3780  
Qy 3781 acttgataaaagcaaaaccccaacttaagcaaaagcggcatcaacagcgggtgtaaaagat 3840  
Db 3781 acttgataaaagcaaaaccccaacttaagcaaaagcggcatcaacagcgggtgtaaaagat 3840  
Qy 3841 taccacattcaatcaggtgagattgcccacaaacagccatgagctgtgacagcgcgcaa 3900  
Db 3841 taccacattcaatcaggtgagattgcccacaaacagccatgagctgtgacagcgcgcaa 3900  
Qy 3901 gatttatgatttaaaaaacccgaacttgaacacaaatcagcagtagctgccaacacagcaca 3960  
Db 3901 gatttatgatttaaaaaacccgaacttgaacacaaatcagcagtagctgccaacacagcaca 3960  
Qy 3961 aaactcattacagaaattctcagtagcagatgaacaaaggttaacttaccgttagtaa 4020  
Db 3961 aaactcattacagaaattctcagtagcagatgaacaaaggttaacttaccgttagtaa 4020

Qy	4021	cccttactccagtttatgacacccctcaagagacctgtatgtctatcacctttgcaggtgaaaa	4080
Db	4021	cccttactccagtttatgacacccctcaagagacctgtatgtctatcacctttgcaggtgaaaa	4080
Qy	4081	cggcattaccaccaaggtataaagtgtggtgtcgtgtggcgatgtgaccaaaccaaaagg	4140
Db	4081	cggcattaccaccaaggtataaagtgtggtgtcgtgtggcgatgtgaccaaaccaaaagg	4140
Qy	4141	cttaaccacgcctaaagtgcacgtgggtataataataatgggcaaaaggcattgtcatgtgacg	4200
Db	4141	cttaaccacgcctaaagtgcacgtgggtataataataatgggcaaaaggcattgtcatgtgacg	4200
Qy	4201	ccaaatgtgtcacaataaccatcacagagactaaagcaaacactctagctaatgttaccacatga	4260
Db	4201	ccaaatgtgtcacaataaccatcacagagactaaagcaaacactctagctaatgttaccacatga	4260
Qy	4261	taaaggtagctacgcaccacagacacggggcaataataatcaaaagacgaagacaacacccg	4320
Db	4261	taaaggtagctacgcaccacagacacggggcaataataatcaaaagacgaagacaacacccg	4320
Qy	4321	tgccgccagcattgttgatgtgtctaagcgcagggtttaacttgcagaagcaatgtgtgaagc	4380
Db	4321	tgccgccagcattgttgatgtgtctaagcgcagggtttaacttgcagaagcaatgtgtgaagc	4380
Qy	4381	ggtgaacttgtctccacttattgacacgcgtcaaacctttccgatggcgaatgccacacccg	4440
Db	4381	ggtgaacttgtctccacttattgacacgcgtcaaacctttccgatggcgaatgccacacccg	4440
Qy	4441	taaggtgacctatgatgacacaaagcaaacacagtaaatgtgtctatgtatgtcatgtgga	4500
Db	4441	taaggtgacctatgatgacacaaagcaaacacagtaaatgtgtctatgtatgtcatgtgga	4500
Qy	4501	tgtataacacattgaagttaaagataaaaaacttggctataaaaaaccacacatgtaccag	4560
Db	4501	tgtataacacattgaagttaaagataaaaaacttggctataaaaaaccacacatgtaccag	4560
Qy	4561	tactggcacagtgctataataatttgccttaagcaatcaagctactggcgatgcgcttgt	4620
Db	4561	tactggcacagtgctataataatttgccttaagcaatcaagctactggcgatgcgcttgt	4620
Qy	4621	caaggccagtgtatgtgtgtctatctataaacaccttatctggcgacatccaaactgccaa	4680
Db	4621	caaggccagtgtatgtgtgtctatctataaacaccttatctggcgacatccaaactgccaa	4680
Qy	4681	aggggcaagccaaagcaaacactcagcaggtcatgtggatgtgatggcaataaaggtcat	4740
Db	4681	aggggcaagccaaagcaaacactcagcaggtcatgtggatgtgatggcaataaaggtcat	4740
Qy	4741	ctatgacagtaccgataacaaagtactatacagcccaaaaatgatggcacagttgataaaaac	4800
Db	4741	ctatgacagtaccgataacaaagtactatacagcccaaaaatgatggcacagttgataaaaac	4800
Qy	4801	caagaagttgccaagacaaactgtgtgcccaagcccaaaccccgatggcacattggc	4860
Db	4801	caagaagttgccaagacaaactgtgtgcccaagcccaaaccccgatggcacattggc	4860
Qy	4861	tcaaatgaaagtcaaatcagctattcaacaaagacaaagtaaatgtatgccaatataaagca	4920
Db	4861	tcaaatgaaagtcaaatcagctattcaacaaagacaaagtaaatgtatgccaatataaagca	4920
Qy	4921	aggoatcaatgaaagacaacgcctttgttaaggaacttgaaaaagccgcttctgtataacaa	4980
Db	4921	aggoatcaatgaaagacaacgcctttgttaaggaacttgaaaaagccgcttctgtataacaa	4980
Qy	4981	aaccataaaacgcgcagtaactgtgggtgatttcaaatgccgttgcgcaaacacgcgtgac	5040
Db	4981	aaccataaaacgcgcagtaactgtgggtgatttcaaatgccgttgcgcaaacacgcgtgac	5040
Qy	5041	ctttgcagggtatcacaggacacacgcgtataaaaactggggcgagactttgaccatcaaaagg	5100
Db	5041	ctttgcagggtatcacaggacacacgcgtataaaaactggggcgagactttgaccatcaaaagg	5100

QY	5101	tgggcaaacagacaccacaaataagctaaocgataataacatctcgggtgtgtagcaggtactga	5160
DB	5101	tgggcaaacagacaccacaaataagctaaocgataataacatcgggtgtgtagcaggtactga	5160
QY	5161	tgggttcaactgtcaaaacttgcacaaagacctaacaactcttaacagcgttaatgcaggtgg	5220
DB	5161	tgggttcaactgtcaaaacttgcacaaagacctaacaactcttaacagcgttaatgcaggtgg	5220
QY	5221	cacaaaaatgatgacaaagcggtgtctttttgtagactcaagcggtaacgccaagcaaa	5280
DB	5221	cacaaaaatgatgacaaagcggtgtctttttgtagactcaagcggtaacgccaagcaaa	5280
QY	5281	cacctctgtctaagtcgcaatgggcttggaactgggtgggaaggtcatcagtaatgtggg	5340
DB	5281	cacctctgtctaagtcgcaatgggcttggaactgggtgggaaggtcatcagtaatgtggg	5340
QY	5341	caaaagcacaaaagatacgcgcgtgcgaatgtacaacagtttaaacgaagtacacgaactt	5400
DB	5341	caaaagcacaaaagatacgcgcgtgcgaatgtacaacagtttaaacgaagtacacgaactt	5400
QY	5401	gttgggtcttggtaatgcttggtaataagctgatacgcgtgacggcgaatcaggtaaacattgcga	5460
DB	5401	gttgggtcttggtaatgcttggtaataagctgatacgcgtgacggcgaatcaggtaaacattgcga	5460
QY	5461	catcaaaaaagacccaaaattcagggttcatactataacgcgactgtcatcaaaagcaggcac	5520
DB	5461	catcaaaaaagacccaaaattcagggttcatactataacgcgactgtcatcaaaagcaggcac	5520
QY	5521	ggtacttggcgttaaggttaataagcgtatccgaaanaacttgcacttgggttatcaagtt	5580
DB	5521	ggtacttggcgttaaggttaataagcgtatccgaaanaacttgggttatcaagtt	5580
QY	5581	ggcgttggataaagacgcgcaacgcgttaacgcgtatttaagcaaatgtttgggtcaaaaccca	5640
DB	5581	ggcgttggataaagacgcgcaacgcgttaacgcgtatttaagcaaatgtttgggtcaaaaccca	5640
QY	5641	aaaagatggcagcaaaaaagccctgtctgcacattataacgcgcgaggtcagacccaacta	5700
DB	5641	aaaagatggcagcaaaaaagccctgtctgcacattataacgcgcgaggtcagacccaacta	5700
QY	5701	tttgaccacaaacccgcgcagaagccattacagaaataaatgaacaaggtatccgcttttt	5760
DB	5701	tttgaccacaaacccgcgcagaagccattacagaaataaatgaacaaggtatccgcttttt	5760
QY	5761	ccatgtcaacgataggcaatacaagagcgtctgggtacaaagggcgttaacgcgcatgtactcaag	5820
DB	5761	ccatgtcaacgataggcaatacaagagcgtctgggtacaaagggcgttaacgcgcatgtactcaag	5820
QY	5821	tgcctcgggcaagcactcagttggcgtatagtttccaggccaaagcgagatgggtgaagccgc	5880
DB	5821	tgcctcaggcaagcactcagttggcgtatagtttccaggccaaagcgagatgggtgaagccgc	5880
QY	5881	cgttgcataaggcagacaaacccaagcaggcaaccaatccatcgccatcgggtgataaagc	5940
DB	5881	cgttgcataaggcagacaaacccaagcaggcaaccaatccatcgccatcgggtgataaagc	5940
QY	5941	acaagccacggggcgatcaaatccatcgccatcgggtacaggcaaatgtgtggtagcaggttaagca	6000
DB	5941	acaagccacggggcgatcaaatccatcgccatcgggtacaggcaaatgtgtggtagcaggttaagca	6000
QY	6001	ctctgtgtccactcggcgaccacaaagcactgttaagggctgataaacagttacagttgtgggttaa	6060
DB	6001	ctctgtgtccactcggcgaccacaaagcactgttaagggctgataaacagttacagttgtgggttaa	6060
QY	6061	taacaaccagtttaccgatgccatcacaacccaatgtcttgggtggggcaataaacatcac	6120
DB	6061	taacaaccagtttaccgatgccatcacaacccaatgtcttgggtggggcaataaacatcac	6120
QY	6121	cgtgaccgaaagtaactcgggttgccttaggttcaaaactctgccatcagtcagtcaggcacaca	6180
DB	6121	cgtgaccgaaagtaactcgggttgccttaggttcaaaactctgccatcagtcagtcaggcacaca	6180
QY	6181	cgcaggtcacacagcccaaaaaattcgcgcgcgcgcgggttacaacccacacacgcagggtgc	6240

```

Db 6181 cgcaggcaacaagcaaaaatctgacggcacagcagggtacaaaccacacagcagggtgc 6240
QY 6241 aacgcgtacggttaaaagcgttctgtgacaaacggcggttgggtgcgtctccgtgggtgc 6300
Db 6241 aacgcgtacggttaaaagcgttctgtgacaaacggcggttgggtgcgtctccgtgggtgc 6300
QY 6301 ctcaagtgctgaacgcgcgtatccaaaatgtggcagcaggtgaggtcagtgccaccagcac 6360
Db 6301 ctcaagtgctgaacgcgcgtatccaaaatgtggcagcaggtgaggtcagtgccaccagcac 6360
QY 6361 cgatgcggtcaatggtgagcaggtgtacaaagccaccacaaagcatttgcacgcgaaccaa 6420
Db 6361 cgatgcggtcaatggtgagcaggtgtacaaagccaccacaaagcatttgcacgcgaaccaa 6420
QY 6421 tgagcttgaccatcgtatccaccacaaacgaaaataagggccaaatgcagggtttcaccgc 6480
Db 6421 tgagcttgaccatcgtatccaccacaaacgaaaataagggccaaatgcagggtttcaccgc 6480
QY 6481 gatggcgatgggtccatgcacaaagcctacattcctcggcagatccatggttaccggggg 6540
Db 6481 gatggcgatgggtccatgcacaaagcctacattcctcggcagatccatggttaccggggg 6540
QY 6541 tattgccaccacaaacggtcaaggtgcggtggcagtggtggactgtcgaagctgtcggataa 6600
Db 6541 tattgccaccacaaacggtcaaggtgcggtggcagtggtggactgtcgaagctgtcggataa 6600
QY 6601 tggtaaatgggtatttaaaatcaatggttcagcgcgatacccaaggccatgtagggcgcc 6660
Db 6601 tggtaaatgggtatttaaaatcaatggttcagcgcgatacccaaggccatgtagggcgcc 6660
QY 6661 agttggcgaggtttcacttttaagcgcataaatacgaagattttactttaaatacaatc 6720
Db 6661 agttggcgaggtttcacttttaagcgcataaatacgaagattttactttaaatacaatc 6720
QY 6721 tcaccatggtgtataaaacagcatcagcatcagtcataattactgctgattgtttt 6780
Db 6721 tcaccatggtgtataaaacagcatcagcatcagtcataattactgctgattgtttt 6780
QY 6781 atcacttaaccattttaccgctcaagtgattctcttccaccatgaccacaaatgcgcatg 6840
Db 6781 atcacttaaccattttaccgctcaagtgattctcttccaccatgaccacaaatgcgcatg 6840
QY 6841 atcatagtgtaaaatcttgataaatttatcaatgtagttgtagatggttaaatt 6900
Db 6841 atcatagtgtaaaatcttgataaatttatcaatgtagttgtagatggttaaatt 6900
QY 6901 gtgcattgacccaaaatgacgatttatcccgaaaatttctgattatgattccgttgac 6960
Db 6901 gtgcattgacccaaaatgacgatttatcccgaaaatttctgattatgattccgttgac 6960
QY 6961 ctgcaggtgcgac 6972
Db 6961 ctgcaggtgcgac 6972

```

## RESULT 2

```

AAF59100
ID AAF59100 standard; DNA; 6973 BP.
XX
AC AAF59100;
XX
DT 24-APR-2001 (first entry)
XX
DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa gene SEQ ID NO:1.
XX
KW Moraxella catarrhalis strain 4223; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection; ds.
XX
OS Moraxella catarrhalis.
XX
PN W0200107619-A1.

```

```

XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-CA00870.
XX
PR 27-JUL-1999; 99US-0361619.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX
WPI: 2001-159722/16.
XX
P-PSDB; AAB59133.
XX
PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
PT useful in protective vaccines and for diagnosis
XX
PS Example 3; Fig 2A-W; 247pp; English.
XX
CC The present invention describes an isolated and purified nucleic acid (I)
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC The 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa gene, which is used
CC in the exemplification of the present invention.
XX
SQ Sequence 6973 BP; 2265 A; 1555 C; 1533 G; 1620 T; 0 other;

```

Query Match 99.8%; Score 6961; DB 22; Length 6973;

Best Local Similarity 100.0%; Pred. NO. 0;

Matches 6972; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ccattgatatggcaggtgtgctgcctgcctgattgagcgatgacacccattgccc 60

Db 1 ccattgatatggcaggtgtgctgcctgcctgattgagcgatgacacccattgccc 60

QY 61 catatctgacgatttgacatgtgatatttaacatgtgacatgatttaacattgttt 120

Db 61 catatctgacgatttgacatgtgatatttaacatgtgacatgatttaacattgttt 120

QY 121 aatactgttgcattatcaacataatttagtaacgcgatttagtaacgatttgaataat 180

Db 121 aatactgttgcattatcaacataatttagtaacgcgatttagtaacgatttgaataat 180

QY 181 cattgcgcccccttattgtatcatatgaatagataattatgattgtatcattgtt 240

Db 181 cattgcgcccccttattgtatcatatgaatagataattatgattgtatcattgtt 240

QY 241 atcagaatggtgctgctatgatgcctacagtgattggttggttaacactctatg 300

Db 241 atcagaatggtgctgctatgatgcctacagtgattggttggttaacactctatg 300

QY 301 atttgatatatttgaactaaatctattgacttaaacaccatattggttaataattagca 360

Db 301 atttgatatatttgaactaaatctattgacttaaacaccatattggttaataattagca 360

QY 361 taatggtgaggtttttgtaaaaataacacatgcgaattgttctactgttactacatgct 420

Db 361 taatggtgaggtttttgtaaaaataacacatgcgaattgttctactgttactacatgct 420

QY 421 tgaatgacgatcccaatcaccagatttccatcaagtgtgtgtttgtacgcaccattta 480

Db 421 tgaatgacgatcccaatcaccagatttccatcaagtgtgtgtttgtacgcaccattta 480



QY 481 cccatattatttcaatcaaatgccttatgtcagcatgtatcatcttttttaaggtaaacac 540  
 Db 481 cccatattatttcaatcaaatgccttatgtcagcatgtatcatcttttttaaggtaaacac 540  
 QY 541 catgaatcacatctataaagtcaatttttaacaaacacagccacatttatggcagtgcc 600  
 Db 541 catgaatcacatctataaagtcaatttttaacaaacacagccacatttatggcagtgcc 600  
 QY 601 agagtacgccaatccacagcac-gggggggggtagctgtgtctacagggcaagtggca 659  
 Db 601 agagtacgccaatccacagcacgggggggggtagctgtgtctacagggcaagtggca 660  
 QY 660 gtgtatgcaactctgagcttgcgcgtattgcgcgtcgcgcgtcgcgtcgcgtcgcgtcgc 719  
 Db 661 gtgtatgcaactctgagcttgcgcgtattgcgcgtcgcgcgtcgcgtcgcgtcgcgtcgc 720  
 QY 720 cgcctagtgagctgtcttatgtctcaaaaaaagataccaaacatatcgcaattggtagac 779  
 Db 721 cgcctagtgagctgtcttatgtctcaaaaaaagataccaaacatatcgcaattggtagac 780  
 QY 780 aaacacagccaagacgctcaggaactgcgaagcgggagcgggtgatcgagccattgctattg 839  
 Db 781 aaacacagccaagacgctcaggaactgcgaagcgggagcgggtgatcgagccattgctattg 840  
 QY 840 gtgaaaatgctaaccacagggcggtcgaagccatcgccatcggttagtagtaataaaactg 899  
 Db 841 gtgaaaatgctaaccacagggcggtcgaagccatcgccatcggttagtagtaataaaactg 900  
 QY 900 tcaatggaagcagtttggataagtagtagtaccgatgtctacgggtcaagagtcacatcgcca 959  
 Db 901 tcaatggaagcagtttggataagtagtagtaccgatgtctacgggtcaagagtcacatcgcca 960  
 QY 960 tcgggtggtagttaaaggctagtggtgatgctcgtatgccatcggttagtagtaccattac 1019  
 Db 961 tcgggtggtagttaaaggctagtggtgatgctcgtatgccatcggttagtagtaccattac 1020  
 QY 1020 atttgcctgtagcagtggttaactcctaaacatccgaagtgactctgtattgaacgactta 1079  
 Db 1021 atttgcctgtagcagtggttaactcctaaacatccgaagtgactctgtattgaacgactta 1080  
 QY 1080 ttaacggccatgcagttataaagaaatacgaagctcaaaagataatgatgtaaataata 1139  
 Db 1081 ttaacggccatgcagttataaagaaatacgaagctcaaaagataatgatgtaaataata 1140  
 QY 1140 gacgcacaacgcgaagcgacagcagtcagtgagtgagccatgtcatatgcacag 1199  
 Db 1141 gacgcacaacgcgaagcgacagcagtcagtcagtgagtgagccatgtcatatgcacag 1200  
 QY 1200 gtcatttttccaaagcctttggtacacgggcaacagctaaagtgcctattccttggcag 1259  
 Db 1201 gtcatttttccaaagcctttggtacacgggcaacagctaaagtgcctattccttggcag 1260  
 QY 1260 tgggtcttgcgcgcacagcgagggccaactcacaatcgctattggttctgtatgcaacat 1319  
 Db 1261 tgggtcttgcgcgcacagcgagggccaactcacaatcgctattggttctgtatgcaacat 1320  
 QY 1320 ctagctcgttgggagcgatagcccttggtaggtagtctgtcgtcagctacagggcagta 1379  
 Db 1321 ctagctcgttgggagcgatagcccttggtaggtagtctgtcgtcagctacagggcagta 1380  
 QY 1380 ttgcccctaggttcaagggttctgttgcactcagagtgataataattctagccggcctata 1439  
 Db 1381 ttgcccctaggttcaagggttctgttgcactcagagtgataataattctagccggcctata 1440  
 QY 1440 caccataatccacagcactagaccccaagtttcaagccaccaataatacgaagcggggtc 1499  
 Db 1441 caccataatccacagcactagaccccaagtttcaagccaccaataatacgaagcggggtc 1500  
 QY 1500 cactttccattggtagtaactctatcaaaacgtaaaaatcaatcaatgctcggtagcagtttta 1559  
 Db 1501 cactttccattggtagtaactctatcaaaacgtaaaaatcaatcaatgctcggtagcagtttta 1560

QY 1560 ataaacccgagtcggtcaatgtggcagcagctagaacggtgggtgaagtgggctaaaggagc 1619  
 Db 1561 ataaacccgagtcggtcaatgtggcagcagctagaacggtgggtgaagtgggctaaaggagc 1620  
 QY 1620 gtgaataattacttttcaggggtgatgataaacagtagtactgacgtataaaatagggttttggaata 1679  
 Db 1621 gtgaataattacttttcaggggtgatgataaacagtagtactgacgtataaaatagggttttggaata 1680  
 QY 1680 ctttaactattaaagggtgtgcagagaccacgcattaaaccgataaaatcgggtgtg 1739  
 Db 1681 ctttaactattaaagggtgtgcagagaccacgcattaaaccgataaaatcgggtgtg 1740  
 QY 1740 taaagaggctgataatagtggtctgaaagttaaactgtctaaacttttaaaacatttta 1799  
 Db 1741 taaagaggctgataatagtggtctgaaagttaaactgtctaaacttttaaaacatttta 1800  
 QY 1800 ctgagggtgaataacacattacattaaatgccacaacacagttaaaggtaggtagtagtagta 1859  
 Db 1801 ctgagggtgaataacacattacattaaatgccacaacacagttaaaggtaggtagtagtagta 1860  
 QY 1860 gtactacagctgaattatttgagtgatagtttaacctttaccagccccaataacagggcagtc 1919  
 Db 1861 gtactacagctgaattatttgagtgatagtttaacctttaccagccccaataacagggcagtc 1920  
 QY 1920 aaagcacaagcaaaacccgtctatggcgttaattggggtagaagtttacttaataatgcagaaa 1979  
 Db 1921 aaagcacaagcaaaacccgtctatggcgttaattggggtagaagtttacttaataatgcagaaa 1980  
 QY 1980 caacagcagcaatcgccactactcgtattaccagagataaaatggcgtttgtctcgcagatg 2039  
 Db 1981 caacagcagcaatcgccactactcgtattaccagagataaaatggcgtttgtctcgcagatg 2040  
 QY 2040 gtgagtgtgataaaacacagccacatatttggataaaaaacacacttaaagtgggttagtg 2099  
 Db 2041 gtgagtgtgataaaacacagccacatatttggataaaaaacacacttaaagtgggttagtg 2100  
 QY 2100 ttgcaattaccatagacaaatggcattgatgcaggttaataaaaaagactcagtaactcttgcca 2159  
 Db 2101 ttgcaattaccatagacaaatggcattgatgcaggttaataaaaaagactcagtaactcttgcca 2160  
 QY 2160 aaggtagcagtgctaacgctggtttacatcgacagctcaaaagcccgccaaagcctacttt 2219  
 Db 2161 aaggtagcagtgctaacgctggtttacatcgacagctcaaaagcccgccaaagcctacttt 2220  
 QY 2220 taaaacgagcgctggcagctcagctgcacacctactgaaataatcagttgatgctaagagtg 2279  
 Db 2221 taaaacgagcgctggcagctcagctgcacacctactgaaataatcagttgatgctaagagtg 2280  
 QY 2280 gcaatgttaccgcccccaacttacaacattggcgtgaaacccacagcagcttaacagtgatg 2339  
 Db 2281 gcaatgttaccgcccccaacttacaacattggcgtgaaacccacagcagcttaacagtgatg 2340  
 QY 2340 gcaatgtatgataaaatttagttaaagggttagtggtacgacacaaatagcttagttaccgccc 2399  
 Db 2341 gcaatgtatgataaaatttagttaaagggttagtggtacgacacaaatagcttagttaccgccc 2400  
 QY 2400 aacatttgcgaagctatcctaaatgaaagtcacacgagcgtgacagtgctctacaaagct 2459  
 Db 2401 aacatttgcgaagctatcctaaatgaaagtcacacgagcgtgacagtgctctacaaagct 2460  
 QY 2460 ttaccgtttaaagaagaagcagatgtagcgcacacgctacccgtggtggttaagaatcacga 2519  
 Db 2461 ttaccgtttaaagaagaagcagatgtagcgcacacgctacccgtggtggttaagaatcacga 2520  
 QY 2520 caaaaaatcccgccgagtcagcatcttaaaactcaaaaggttaaaacggtcttaacggttg 2579  
 Db 2521 caaaaaatcccgccgagtcagcatcttaaaactcaaaaggttaaaacggtcttaacggttg 2580  
 QY 2580 ctacaaaaaaagatggtacggtttacotttgggttggtgacgaagatagcgtctgacacatg 2639  
 Db 2581 ctacaaaaaaagatggtacggtttacotttgggttggtgacgaagatagcgtctgacacatg 2640  
 QY 2640 gcaaaagcacccttaaaacacgagtggttgaagtataccacgaacaaatccaaag 2699

||||| 2641 gcaaaagcacccataaacaacgagtgccctgactggttaagatataccaacgaacaaatccaag 2700  
QY tcggtgctaattggtcattataattactaattgtaattgtaattgtaattgtaattgtaattgtaattg 2759  
||||| 2701 tcggtgctaattggtcattataattactaattgtaattgtaattgtaattgtaattgtaattg 2760  
QY 2760 caaataccgctcgcatcaccagagataaaattggctttgcttgcttgatgagtggtgcagttg 2819  
Db 2761 caaataccgctcgcatcaccagagataaaattggctttgcttgcttgatgagtggtgcagttg 2820  
QY 2820 atacaaacaacattctcttgatcaagacaagctacaagttggcaattgtaattgtaattgtaattg 2879  
Db 2821 atacaaacaacattctcttgatcaagacaagctacaagttggcaattgtaattgtaattgtaattg 2880  
QY 2880 aactggtcattaaacgaggtggttaaacccatcacaggtggtccccaacactgctctagca 2939  
Db 2881 aactggtcattaaacgaggtggttaaacccatcacaggtggtccccaacactgctctagca 2940  
QY 2940 ttgcgcatcaaatgtagcgcgaacatagactgggcaatacaatcccaagacaagaacaaat 2999  
Db 2941 ttgcgcatcaaatgtagcgcgaacatagactgggcaatacaatcccaagacaagaacaaat 3000  
QY 3000 ccaacgctgcagcatttaattgatatataataacagagctttaaccttaaaaaataataaca 3059  
Db 3001 ccaacgctgcagcatttaattgatatataataacagagctttaaccttaaaaaataataaca 3060  
QY 3060 acccattgactttgtctccacttatgacattgttgactttgccaatggcaatggccaacca 3119  
Db 3061 acccattgactttgtctccacttatgacattgttgactttgccaatggcaatggccaacca 3120  
QY 3120 ccgccaagtaaccattgataccgcttaacaaacaaacagtaaaagtggatgagtggaatg 3179  
Db 3121 ccgccaagtaaccattgataccgcttaacaaacaaacagtaaaagtggatgagtggaatg 3180  
QY 3180 tggatgatacaaccattcatcctaacaggaactgatgacaataaaaaacttggcgtaaaaa 3239  
Db 3181 tggatgatacaaccattcatcctaacaggaactgatgacaataaaaaacttggcgtaaaaa 3240  
QY 3240 ccaccaactgaacaaacaaagtgcttaattgtaatacagcaactaaactttaattgtaact 3299  
Db 3241 ccaccaactgaacaaacaaagtgcttaattgtaatacagcaactaaactttaattgtaact 3300  
QY 3300 ctagtgtgaagtgccttgtttaacgccaaagacatcgcgcgaataatcctaaccacctag 3359  
Db 3301 ctagtgtgaagtgccttgtttaacgccaaagacatcgcgcgaataatcctaaccacctag 3360  
QY 3360 ccaaggaaattcacaccacaaaggacagcagacacgccttacaacacottttaccggtta 3419  
Db 3361 ccaaggaaattcacaccacaaaggacagcagacacgccttacaacacottttaccggtta 3420  
QY 3420 aaaaggtagatgaaaaataataatgctgatgacgcgaacgcacatcacccgtgggtcaaaaga 3479  
Db 3421 aaaaggtagatgaaaaataataatgctgatgacgcgaacgcacatcacccgtgggtcaaaaga 3480  
QY 3480 acgcaataatcaagtcaacacccctaaactcaaaagtgaacacggtcttaataataaaa 3539  
Db 3481 acgcaataatcaagtcaacacccctaaactcaaaagtgaacacggtcttaataataaaa 3540  
QY 3540 ccgacaaaaattggtacggttacctttgcatataacacccaagcgtcttaaaagccggca 3599  
Db 3541 ccgacaaaaattggtacggttacctttgcatataacacccaagcgtcttaaaagccggca 3600  
QY 3600 aaagcacccctaaaacgacggtggcttgcattataaaaaacccactggtagcggaacaaatcc 3659  
Db 3601 aaagcacccctaaaacgacggtggcttgcattataaaaaacccactggtagcggaacaaatcc 3660  
QY 3660 aagtcggtgctgataaggcgtgaagtgttccaaaggttaataataatggtggtgagtgctg 3719  
Db 3661 aagtcggtgctgataaggcgtgaagtgttccaaaggttaataataatggtggtgagtgctg 3720  
QY 3720 gcatgtggtgcacaaactcgcatcaccagagataaaattggctttacttgggactaatggct 3779  
|||||

Db 3721 gcatgtggtgcacaaactcgcatcaccagagatgaaattggctttacttgggactaatggct 3780  
QY 3780 cacttgataaaagcaaacccacccacttaagcaaacgacggtcattaaacgaggtgtgtaaaaaga 3839  
Db 3781 cacttgataaaagcaaacccacccacttaagcaaacgacggtcattaaacgaggtgtgtaaaaaga 3840  
QY 3840 ttaccaacattcaatcagggtgagattggcccaaaacagccatgagtctgtgacagggcgca 3899  
Db 3841 ttaccaacattcaatcagggtgagattggcccaaaacagccatgagtctgtgacagggcgca 3900  
QY 3900 agatttatgatttaaaacccgaaacttgaacaaaataacagcagctactgcgaacacagcac 3959  
Db 3901 agatttatgatttaaaacccgaaacttgaacaaaataacagcagctactgcgaacacagcac 3960  
QY 3960 aaaactcattacacgaattctcagtagcagatgaacaaaggttaataacttttacggttagta 4019  
Db 3961 aaaactcattacacgaattctcagtagcagatgaacaaaggttaataacttttacggttagta 4020  
QY 4020 acccttactcagttatgacacctcaagacctctgatgtcatcacctttgcaggtgaaa 4079  
Db 4021 acccttactcagttatgacacctcaagacctctgatgtcatcacctttgcaggtgaaa 4080  
QY 4080 acggcattaccacaaaggttaataaaggtgtgtgctgtgtggcattgacacaaacacaaag 4139  
Db 4081 acggcattaccacaaaggttaataaaggtgtgtgctgtgtggcattgacacaaacacaaag 4140  
QY 4140 gcttaacacgctcaagctgacccgtgggttaataataatggcaagggcattgtcattgaca 4199  
Db 4141 gcttaacacgctcaagctgacccgtgggttaataataatggcaagggcattgtcattgaca 4200  
QY 4200 gccaaaaatggttcaaaaaataccatcacaggaacttaagcaacactctagctaatgttaccacatg 4259  
Db 4201 gccaaaaatggttcaaaaaataccatcacaggaacttaagcaacactctagctaatgttaccacatg 4260  
QY 4260 ataaagtagcgttacgcacacacagacagggccaataatacaaaagagaagacaaaacccc 4319  
Db 4261 ataaagtagcgttacgcacacacagacagggccaataatacaaaagagaagacaaaacccc 4320  
QY 4320 gtgcgcgcagcattgtgtgcttaagcgcagcgtttaaacttgcgaaggaactggtgaag 4379  
Db 4321 gtgcgcgcagcattgtgtgcttaagcgcagcgtttaaacttgcgaaggaactggtgaag 4380  
QY 4380 cggttgactttgtctccacttatgacacccgttcaactttgcgatggcaatgcccaccacgg 4439  
Db 4381 cggttgactttgtctccacttatgacacccgttcaactttgcgatggcaatgcccaccacgg 4440  
QY 4440 ctaaggtgacctatgtagacacaaagcaaacccagtaaaagtgtgtctatgattcgaatgtgg 4499  
Db 4441 ctaaggtgacctatgtagacacaaagcaaacccagtaaaagtgtgtctatgattcgaatgtgg 4500  
QY 4500 atgatacaacccattgaagttaaagataaaaaacttggcgttaaaaaacccacacattgacca 4559  
Db 4501 atgatacaacccattgaagttaaagataaaaaacttggcgttaaaaaacccacacattgacca 4560  
QY 4560 gtactggcacaggtgcttaataaaatttgccttaagcaatcaagctacttggcgatgcgcttg 4619  
Db 4561 gtactggcacaggtgcttaataaaatttgccttaagcaatcaagctacttggcgatgcgcttg 4620  
QY 4620 tcaagggcaggtgactgtgctcatcaaaccttatctggcgacatccaaactgcca 4679  
Db 4621 tcaagggcaggtgactgtgctcatcaaaccttatctggcgacatccaaactgcca 4680  
QY 4680 aaggggcaagccaaagcaacaaactcagcaggtctgtggaatggtgacgaatgaaggttca 4739  
Db 4681 aaggggcaagccaaagcaacaaactcagcaggtctgtggaatggtgacgaatgaaggttca 4740  
QY 4740 tctatgacagtacogatacaaaagtactatcaagccaaaaatgtaggcacagttgataaaa 4799  
Db 4741 tctatgacagtacogatacaaaagtactatcaagccaaaaatgtaggcacagttgataaaa 4800  
QY 4800 ccaaaagagttgccaagacaaactggtcgcccaagcccaaacccagatggcacattgg 4859  
Db 4801 ccaaaagagttgccaagacaaactggtcgcccaagcccaaacccagatggcacattgg 4860



## RESULT 3

```

AAT38740
ID   AAT38740 standard; DNA; 6971 BP.
XX   AC
XX   AAT38740;
XX   DT
XX   25-JAN-1997 (first entry)
XX   DE
XX   Moraxella outer membrane protein gene.
XX   KW
XX   Outer membrane protein; OMP; immunogen; vaccine; otitis media;
XX   diagnosis; ss.
XX   OS
XX   Mycobacterium catarrhalis strain 4223.
XX   FH
XX   Key
XX   CDS
XX   706..6684
XX   FT
XX   /*tag= a
XX   PN
XX   W09634960-A1.
XX   PD
XX   07-NOV-1996.
XX   PF
XX   29-APR-1996; 96WO-CA00264.
XX   PR
XX   26-MAR-1996; 96US-0621944.
XX   PR
XX   01-MAY-1995; 95US-0431718.
XX   PR
XX   07-JUN-1995; 95US-0478370.
XX   PA
XX   (CONN-) CONNAUGHT LAB LTD.
XX   PI
XX   Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;
XX   WPI; 1996-506162/50.
XX   DR
XX   P-PSDB; AAW04505.
XX   PT
XX   Moraxella outer membrane protein - useful as immunogen in protective
XX   vaccine and for diagnosis
XX   PS
XX   Claim 14; Fig 6; 109pp; English.
XX   CC
XX   A genomic DNA fragment (738740) of Moraxella catarrhalis otitis
XX   media strain 4223 includes the coding region for a 1992-amino
XX   acid protein (W04505) identified as an approx. 200 kDa outer
XX   membrane protein (OMP). The DNA was isolated from a strain 4223
XX   genomic library in phage lambda EMBL3 by screening with an
XX   anti-200 kDa protein guinea pig antiserum. The gene can be
XX   used for the recombinant expression of the OMP (for use in
XX   vaccines), for the prepn. of hybridisation probes, or may be
XX   incorporated into a live vector for use in direct immunisation.
XX   SQ
XX   Sequence 6971 BP; 2265 A; 1553 C; 1533 G; 1620 T; 0 other;

Query Match          99.98; Score 6959; DB 17; Length 6971;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6970; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY   3 atggatatggcagggtgtgctgcctgcctgtatgtatggcagatgacacccattggcccca 62
DB   1 atggatatggcagggtgtgctgcctgcctgcctgtatgtatggcagatgacacccattggcccca 60

QY   63 tatctgtacgatttgacatgtgatattgaatttaacatgtgacatgatttaacattgtttaa 122
DB   61 tatctgtacgatttgacatgtgatattgaatttaacatgtgacatgatttaacattgtttaa 120

QY   123 tactgttgccatcattaccataattttagtaacgcatttagtaacgcatttggtaaaaaatca 182
DB   121 tactgttgccatcattaccataattttagtaacgcatttagtaacgcatttggtaaaaaatca 180

QY   183 ttgcgccccctttatgtatcatatgaatagaattatgattgtatctgattatttat 242
DB   181 ttgcgccccctttatgtatcatatgaatagaattatgattgtatctgattatttat 240

```

```

QY   243 cagaatggtgatgctatgatgatgctacgagttgatttgggttaataactactctatgat 302
DB   241 cagaatggtgatgctatgatgatgctacgagttgatttgggttaataactactctatgat 300

QY   303 ttgatataattttgaaactaatctattgacttaaatcaccatataatggttaataattagcata 362
DB   301 ttgatataattttgaaactaatctattgacttaaatcaccatataatggttaataattagcata 360

QY   363 atggtaggctttttgtaaaaaatcacatcgcaattgttctactgttactaccatgcttg 422
DB   361 atggtaggctttttgtaaaaaatcacatcgcaattgttctactgttactaccatgcttg 420

QY   423 aatgacgatcccaatcaccagatttcattcaagtgtatgtttgttatatcgaccattttacc 482
DB   421 aatgacgatcccaatcaccagatttcattcaagtgtatgtttgttatatcgaccattttacc 480

QY   483 ctaattatttcaatacaaatgctatgtcagcatgtatcatattttttaaggtaaacaccca 542
DB   481 ctaattatttcaatacaaatgctatgtcagcatgtatcatattttttaaggtaaacaccca 540

QY   543 tgaatcacatctataaagtcatctttaaaaaagccacagccacattttatgcagtggcag 602
DB   541 tgaatcacatctataaagtcatctttaaaaaagccacagccacattttatgcagtggcag 600

QY   603 agtacgcaaaatccacacagcac-99ggggggggtagctgtgtacagggcaagttggcagt 661
DB   601 agtacgcaaaatccacacagcacg99ggggggggtagctgtgtacagggcaagttggcagt 660

QY   662 gtatgcacttgagctttgcccgtattgcccgtattgcccgtctgctgctcgtgctgctgctg 721
DB   661 gtatgcacttgagctttgcccgtattgcccgtattgcccgtctgctgctcgtgctgctgctg 720

QY   722 ctacgtggcagtgcttatgtctcaaaaaaagatcaccaaacatatacgcaattggtgaacaa 781
DB   721 ctacgtggcagtgcttatgtctcaaaaaaagatcaccaaacatatacgcaattggtgaacaa 780

QY   782 aaccagccaagcagctcagggcactgccaagcgacgggtgacgagcattgctattggt 841
DB   781 aaccagccaagcagctcagggcactgccaagcgacgggtgacgagcattgctattggt 840

QY   842 gaaatgctaacgcacagggcggtgcaagccatcgccatcgctagtagtaataaaactgtc 901
DB   841 gaaatgctaacgcacagggcggtgcaagccatcgccatcgctagtagtaataaaactgtc 900

QY   902 aatggaagcagcttttgataaagataggtaccgctacggtcaggtcgaagatccatcgccatc 961
DB   901 aatggaagcagcttttgataaagataggtaccgctacggtcaggtcgaagatccatcgccatc 960

QY   962 ggtggtgatgtaaaaggctagtggtgatgctcgtcattgcccattcggtagtgacttaact 1021
DB   961 ggtggtgatgtaaaaggctagtggtgatgctcgtcattgcccattcggtagtgacttaact 1020

QY   1022 ttgcttgatcagcagtgttaactccttaaacatccgaaaagctactctgattaaacgatttatt 1081
DB   1021 ttgcttgatcagcagtgttaactccttaaacatccgaaaagctactctgattaaacgatttatt 1080

QY   1082 aacggccatgcagtatataaagaataacgaagctcaaaaggaataatgattgtaaaatataga 1141
DB   1081 aacggccatgcagtatataaagaataacgaagctcaaaaggaataatgattgtaaaatataga 1140

QY   1142 cgcaacacgcgaagcggaacacgcagctactgcagtggtgagccatgctatgcacaggt 1201
DB   1141 cgcaacacgcgaagcggaacacgcagctactgcagtggtgagccatgctatgcacaggt 1200

QY   1202 cattttccaacgccttttgggtacacgggcaacagctaaaaagtgccatttccttggcagtg 1261
DB   1201 cattttccaacgccttttgggtacacgggcaacagctaaaaagtgccatttccttggcagtg 1260

QY   1262 ggtcttccgcacacgcagcgaggcccaatctacaatcgctattggttctgctgacacatct 1321
DB   1261 ggtcttccgcacacgcagcgaggcccaatctacaatcgctattggttctgctgacacatct 1320

```



Db	3481	gcaataataatcaangtcacacccctaacactcaagdtgaaacggtctttaataataaac	3541
Qy	3542	gacaaaaatgttacggttacctttggcattaacaccacaagcgtcttaaaagccgcgcaaa	3601
Db	3541	gacaaaaatgttacggttacctttggcattaacaccacaagcgtcttaaaagccgcgcaaa	3600
Qy	3602	agacccctaaaagcagcgtgtgctgtctattaaaaaccccactggtagcgacaaatcccaa	3661
Db	3601	agacccctaaaagcagcgtgtgctgtctattaaaaaccccactggtagcgacaaatcccaa	3660
Qy	3662	gtcggtcctgatggcgtgaaagtcttccaaagcttaataataatggttgttagtgtgcgc	3721
Db	3661	gtcgggtcctgatggcgtgaaagtcttccaaagcttaataataatggttgttagtgtgcgc	3720
Qy	3722	attgatggcacaaactcgctattaccagagatgaattggctttactgggactaatggctca	3781
Db	3721	attgatggcacaaactcgctattaccagagatgaattggctttactgggactaatggctca	3780
Qy	3782	cttgataaaagcaaaccccaactaagcaagaacggcattaacgcaggtgtgtaaaaagatt	3841
Db	3781	cttgataaaagcaaaccccaactaagcaagaacggcattaacgcaggtgtgtaaaaagatt	3840
Qy	3842	acaaacattcaatcagggtgagattgcccaaaacagccatgatgctgtgacagcgcgcgaag	3901
Db	3841	acaaacattcaatcagggtgagattgcccaaaacagccatgatgctgtgacagcgcgcgaag	3900
Qy	3902	atttatgatttaaaaacccgaacttgaaaacaaaaatcagcagctactgccaaaacagcacaa	3961
Db	3901	atttatgatttaaaaacccgaacttgaaaacaaaaatcagcagctactgccaaaacagcacaa	3960
Qy	3962	aactcattacagaaattctcagtagcagatgaacaaggttaataactttacggttagtaac	4021
Db	3961	aactcattacagaaattctcagtagcagatgaacaaggttaataactttacggttagtaac	4020
Qy	4022	ccctactccagttatgacacctcaagacctctgatgtcatcacccttgcaggtgaaac	4081
Db	4021	ccctactccagttatgacacctcaagacctctgatgtcatcacccttgcaggtgaaac	4080
Qy	4082	ggcattaccaccaaggttaataaaggctggctggcgttgggcattgaccaaaaccaaggc	4141
Db	4081	ggcattaccaccaaggttaataaaggctggctggcgttgggcattgaccaaaaccaaggc	4140
Qy	4142	ttaaccacgcctaaagctgaccctggggttaataatggcaaggcattgtcattgacgc	4201
Db	4141	ttaaccacgcctaaagctgaccctggggttaataatggcaaggcattgtcattgacgc	4200
Qy	4202	caaaaatggtcaaaaatccatcacaggactaagcaacaccttagtctaattgtaccaatgat	4261
Db	4201	caaaaatggtcaaaaatccatcacaggactaagcaacaccttagtctaattgtaccaatgat	4260
Qy	4262	aaagggtagcgtacgcacccacagacagcggcgaataataatcaaaagcgaagcgaaccccg	4321
Db	4261	aaagggtagcgtacgcacccacagacagcggcgaataataatcaaaagcgaagcgaaccccg	4320
Qy	4322	ggcgcgcagcatgttgtatgtctgaagcgcagcgttcaacttgcgaagcgaatggtgaagcg	4381
Db	4321	ggcgcgcagcatgttgtatgtctgaagcgcagcgttcaacttgcgaagcgaatggtgaagcg	4380
Qy	4382	gttgactttgtctccacttatgacacggtcaacttgcgatggcgaatgcaaccacgcgt	4441
Db	4381	gttgactttgtctccacttatgacacggtcaacttgcgatggcgaatgcaaccacgcgt	4440
Qy	4442	aaagtgacctatgatgacacaagcaaaacccagtaaaagtgtctctatgatgcaatgggat	4501
Db	4441	aaagtgacctatgatgacacaagcaaaacccagtaaaagtgtctctatgatgcaatgggat	4500
Qy	4502	gatacaacattgaagttaaagataaaaaacttggcttaaaacccaacattgaccagct	4561
Db	4501	gatacaacattgaagttaaagataaaaaacttggcttaaaacccaacattgaccagct	4560
Qy	4562	actggcacaggtgtctaataaaatttgccttaagcaataacgactctggcgatggcgttgtc	4621
Db	4561	actggcacaggtgtctaataaaatttgccttaagcaataacgactctggcgatggcgttgtc	4620

QY	4622	aaggccagtgatctgcttgctcatctaaacaccttatcttggcgcatccaactgcgcaaa	4680
Db	4621	aaggccagtgatctgcttgctcatctaaacaccttatctggcgatccaaacgtccaaa	4680
QY	4682	ggggcaagccaagcgaaactcagcaggtatgtggatgctgcatggaataaagttcatc	4741
Db	4681	ggggcaagccaagcgaaactcagcaggtatgtggaagtgaaggcaataaagttcatc	4740
QY	4742	tatgacagtcaccgatacaacagttactatcaagccaaaaatgtagcacagttgataaaaacc	4801
Db	4741	tatgacagtcaccgatacaacagttactcaagccaaaaatgtagcacagttgataaaaacc	4800
QY	4802	aaagaattgccaaaagacaactggtgcgccaaagcccaacccccagatggcacattggct	4861
Db	4801	aaagaattgccaaaagacaactggtgcgccaaagcccaacccccagatggcacattggct	4860
QY	4862	caaatgaattcaaaatcagtcattacaagaacaaagtaaatgtagtccaaataaaagcaa	4921
Db	4861	caaatgaattcaaaatcagtcattacaagaacaaagtaaatgtagtccaaataaaagcaa	4920
QY	4922	ggcatcaatgaagacaacgccttgtttaaagagactgaaaaagccgctcttgataaacaaa	4981
Db	4921	ggcatcaatgaagacaacgccttgtttaaagagactgaaaaagccgctcttgataaacaaa	4980
QY	4982	accaaaacgcgcagtaactggtggtatttaaattgcgcttgcccaaacacccgtgacc	5041
Db	4981	accaaaacgcgcagtaactggtggtatttaaattgcgcttgcccaaacacccgtgacc	5040
QY	5042	tttcaggggatacaggcacaacgcgtataaaaactggcgagactttgaccatcaaaaggt	5101
Db	5041	tttcaggggatacaggcacaacgcgtataaaaactggcgagactttgaccatcaaaaggt	5100
QY	5102	gggcaacagacaccaataagctlaaacgcgataataacatcggtgtggtgacgagttactgat	5161
Db	5101	gggcaacagacaccaataagctlaaacgcgataataacatcggtgtggtgacgagttactgat	5160
QY	5162	ggcttcactgtcaacttgcgaagacctaaccatcttaacagcgttaatgcaaggtggc	5221
Db	5161	ggcttcactgtcaacttgcgaagacctaaccatcttaacagcgttaatgcaaggtggc	5220
QY	5222	accaaaattgatgacaaaggcgtgtctttgtagactcaagcgggtcaagccaaagcaaac	5281
Db	5221	accaaaattgatgacaaaggcgtgtctttgtagactcaagcgggtcaagccaaagcaaac	5280
QY	5282	accctgtcttaagtgcgaatggcttggacctgggtggcgaagttcatcagtaaatgtgggc	5341
Db	5281	accctgtcttaagtgcgaatggcttggacctgggtggcgaagttcatcagtaaatgtgggc	5340
QY	5342	aaaggcaaaaagataccgcgcgtgccatgtacaacagttaaacgaagtcgcgaacttg	5401
Db	5341	aaaggcaaaaagataccgcgcgtgccatgtacaacagttaaacgaagtcgcgaacttg	5400
QY	5402	tggggtttggtaatgctggtgaatgataacgctgacggcgaatcagggtataaacattgcgcac	5461
Db	5401	tggggtttggtaatgctggtgaatgataacgctgacggcgaatcagggtataaacattgcgcac	5460
QY	5462	atcaaaaaaacccaaattcagtttcattcatctaaccgcactgtctaatcaaaagcagcag	5521
Db	5461	atcaaaaaaacccaaattcagtttcattcatctaaccgcactgtctaatcaaaagcagcag	5520
QY	5522	gtacttggcgttaaggttaataacgataccgaaaaaacttgcacattgggtatacaagt	5581
Db	5521	gtacttggcgttaaggttaataacgataccgaaaaaacttgcacattgggtatacaagt	5580
QY	5582	ggcgtgataaagacggcaacgctaacggcgatttaagcaattgtttgggtcaaaaccccaa	5641
Db	5581	ggcgtgataaagacggcaacgctaacggcgatttaagcaattgtttgggtcaaaaccccaa	5640
QY	5642	aaagatggcagcaaaaaagccctgctgcgaacttataacgcgcaggttcagaccaactat	5701
Db	5641	aaagatggcagcaaaaaagccctgctgcgaacttataacgcgcaggttcagaccaactat	5700



5702 ttgaccacaaccccgagagccattgacagaataaaatgaacaagggtatcgcttcttc 5761  
|||||  
5701 ttgaccacaaccccgagagccattgacagaataaaatgaacaagggtatcgcttcttc 5760  
|||||  
5762 catgtcaacgatggcaatcaagagcctgtggtacaaaggcgtaacggcattgactcaagt 5821  
|||||  
5761 catgtcaacgatggcaatcaagagcctgtggtacaaaggcgtaacggcattgactcaagt 5820  
|||||  
5822 gctcaggcaagcactcagtgccgatagtttccaggccaagcgagatggtgaagccccc 5881  
|||||  
5821 gctcaggcaagcactcagtgccgatagtttccaggccaagcgagatggtgaagccccc 5880  
|||||  
5882 gttgcatagggcagacaaccccaagcaggtcaacccaatccatgccatcggtgataacgca 5941  
|||||  
5881 gttgcatagggcagacaaccccaagcaggtcaacccaatccatgccatcggtgataacgca 5940  
|||||  
5942 caagccacggggcgatcaatccatcgccatcggttacaggccaatgtggtagcaggttaagcac 6001  
|||||  
5941 caagccacggggcgatcaatccatcgccatcggttacaggccaatgtggtagcaggttaagcac 6000  
|||||  
6002 tctgtgcatcggtcgacccaagcactgtaaggctgataaacagttacagttcaggtgtaat 6061  
|||||  
6001 tctgtgcatcggtcgacccaagcactgtaaggctgataaacagttacagttcaggtgtaat 6060  
|||||  
6062 aacaaacagtttaccgatgccactcaaacccgatgtcttgggtgggccaataacatcaccc 6121  
|||||  
6061 aacaaacagtttaccgatgccactcaaacccgatgtcttgggtgggccaataacatcaccc 6120  
|||||  
6122 gtgacggaagtaactcggttgcttagtttcaaacctctgccatcagtcagtcagccacacac 6181  
|||||  
6121 gtgacggaagtaactcggttgcttagtttcaaacctctgccatcagtcagtcagccacacac 6180  
|||||  
6182 gcaggcacacaagcccaaaaaatctgcgcacagcaggttacaacccaccacagcaggtgca 6241  
|||||  
6181 gcaggcacacaagcccaaaaaatctgcgcacagcaggttacaacccaccacagcaggtgca 6240  
|||||  
6242 accggtacggttaaaagcttctgtgacaaaacggtggtggtgcttccgtggtggtgcc 6301  
|||||  
6241 accggtacggttaaaagcttctgtgacaaaacggtggtggtgcttccgtggtggtgcc 6300  
|||||  
6302 tcagggtgtaacgcgctatccaaaatgtgcagcaggtgaggtcagtcagtcacccagcaccc 6361  
|||||  
6301 tcagggtgtaacgcgctatccaaaatgtgcagcaggtgaggtcagtcagtcacccagcaccc 6360  
|||||  
6362 gatcggtgtaacgtgtagcaggttgcacaaagccacccaagcattgccaaacgcaaccaat 6421  
|||||  
6361 gatcggtgtaacgtgtagcaggttgcacaaagccacccaagcattgccaaacgcaaccaat 6420  
|||||  
6422 gacttgaccatgtagtccaaacaaacgaataaaggccaatgcaggtatttcacagcg 6481  
|||||  
6421 gacttgaccatgtagtccaaacaaacgaataaaggccaatgcaggtatttcacagcg 6480  
|||||  
6482 atggcgatggctccatgcacaaagctacattctcggcagatccatggttacccgggggt 6541  
|||||  
6481 atggcgatggctccatgcacaaagctacattctcggcagatccatggttacccgggggt 6540  
|||||  
6542 atgtccacccacaacggttcaaggtgcggtggcggtgaggtgcgaagctgcgagataat 6601  
|||||  
6541 atgtccacccacaacggttcaaggtgcggtggcggtgaggtgcgaagctgcgagataat 6600  
|||||  
6602 ggtcaatggtatttaaaatcaatggttcagccgatacccaagccatgtaggggcgcgca 6661  
|||||  
6601 ggtcaatggtatttaaaatcaatggttcagccgatacccaagccatgtaggggcgcgca 6660  
|||||  
6662 gttggtcaggttttccacttttaagccataaatacgaagattttacttaaaaaatcaatct 6721  
|||||  
6661 gttggtcaggttttccacttttaagccataaatacgaagattttacttaaaaaatcaatct 6720  
|||||  
6722 caccatggtgataaaacagcatcagcatcagtcattactgactgactgatttttta 6781  
|||||  
6721 caccatggtgataaaacagcatcagcatcagtcattactgactgactgatttttta 6780  
|||||  
6782 tcacttaaacacattttaccgctcaagtgattctcttcttccaccatgaaccaatcgccattga 6841

|||||  
6781 tcacttaaacacattttaccgctcaagtgattctcttccaccatgacaaaatcgccattga 6840  
|||||  
6842 tcataaggtaaaacttattgagtaaaattttatcaatgtagtttagatatggttaaaattg 6901  
|||||  
6841 tcataaggtaaaacttattgagtaaaattttatcaatgtagtttagatatggttaaaattg 6900  
|||||  
6902 tgccattgaccaaataatgaccgatttatcccgaaaatttctgattatgacccgttgacc 6961  
|||||  
6901 tgccattgaccaaataatgaccgatttatcccgaaaatttctgattatgacccgttgacc 6960  
|||||  
6962 tgcagggtcgac 6972  
|||||  
6961 tgcagggtcgac 6971

## RESULT 4

AAF59106

ID AAF59106 standard; DNA; 6259 BP.

XX

AC AAF59106;

DT 24-APR-2001 (first entry)

XX

DE M. catarrhalis M56 200kDa gene in pKS348 SEQ ID NO:12.

XX

KW Moraxella catarrhalis strain Q8; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KW otitis media; detection; ds.

XX

OS Moraxella catarrhalis.

XX

PN WO200107619-A1.

XX

PD 01-FEB-2001.

XX

PF 26-JUL-2000; 2000WO-CA00870.

XX

PR 27-JUL-1999; 99US-0361619.

XX

(CONN-) CONNAUGHT LAB LTD.

XX

PI Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX

DR WPI: 2001-159722/16.

XX

DR P-PSDB; AAB69127.

XX

PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

XX useful in protective vaccines and for diagnosis

XX

PS Claim 1; Fig 8A-V; 247pp; English.

XX

The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II), and its truncated versions, are used as immunogenic compositions and vaccines to protect against M. catarrhalis infections, particularly otitis media in humans. (II) is also used as antigen in immunoassays for detecting specific antibodies (Ab), and to generate Ab. (I) are used for recombinant production of (II) and its fragments are used as probes for identifying/cloning 200 kDa protein genes from other strains, and for diagnostic detection of M. catarrhalis. (I) makes possible production of large amount of recombinant immunogens. Expression of truncated versions of (II) reduces toxicity of the protein towards the Escherichia coli host. The present sequence represents the M. catarrhalis M56 200kDa gene in pKS348, which is given in the exemplification of the present invention.

XX

SQ Sequence 6259 BP; 2067 A; 1414 C; 1393 G; 1385 T; 0 other;

Query Match

89.8%; Score 6258; DB 22; Length 6259;

Best Local Similarity

100.0%; Pred. No. 0;



Matches 6258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 708 tgatcggtgcaacgctcagtggcagtgcttatgtctcaaaaaaagatacacaacatctcg 767  
Db 2 tgatcgggtgcaacgctcagtggcagtgcttatgtctcaaaaaaagatacacaacatctcg 61  
Qy 768 caattgtgtaacaaacacagcaagcagctcagcactgccaagcgccagcggtgaatcgag 827  
Db 62 caattgtgtaacaaacacagcaagcagctcagcactgccaagcgccagcggtgaatcgag 121  
Qy 828 ccattgttattggtgtaaaatgctaacgcacagggcggtcaagccatccgcctcgtgtagta 887  
Db 122 ccattgttattggtgtaaaatgctaacgcacagggcggtcaagccatccgcctcgtgtagta 181  
Qy 888 gtaataaaactgtcaattggaagcagtttggataagataggtaccgatagtctcagggccaag 947  
Db 182 gtaataaaactgtcaattggaagcagtttggataagataggtaccgatagtctcagggccaag 241  
Qy 948 agtccatcgccatcggtggtgtagtataaggttagtgatgctcgaattgccatcgta 1007  
Db 242 agtccatcgccatcggtggtgtagtataaggttagtgatgctcgaattgccatcgta 301  
Qy 1008 gtgatgacttacatttctgtgatcagcatggttaactccttaaacatccgaaaggttactctga 1067  
Db 302 gtgatgacttacatttctgtgatcagcatggttaactccttaaacatccgaaaggttactctga 361  
Qy 1068 ttaacgatcttattaaacggccatcgatgattataaagaataacgaagcttcaaaagataatg 1127  
Db 362 ttaacgatcttattaaacggccatcgatgattataaagaataacgaagcttcaaaagataatg 421  
Qy 1128 atgtataaatatagacgcacacccgcaagcgacacgcagctactgcagtgaggagccatgt 1187  
Db 422 atgtataaatatagacgcacacccgcaagcgacacgcagctactgcagtgaggagccatgt 481  
Qy 1188 catatgcacagggctcattttcccaacgctttggtacacggcgcaacagctaaaaagtgacct 1247  
Db 482 catatgcacagggctcattttcccaacgctttggtacacggcgcaacagctaaaaagtgacct 541  
Qy 1248 attccttggaagtgagggttctgcgccacagccgagggcccaatctacaatcgctatttggtt 1307  
Db 542 attccttggaagtgagggttctgcgccacagccgagggcccaatctacaatcgctatttggtt 601  
Qy 1308 ctgatgcaacatctagctcgttggagcgatagcccttggcgaggttactcgtctcagc 1367  
Db 602 ctgatgcaacatctagctcgttggagcgatagcccttggcgaggttactcgtctcagc 661  
Qy 1368 tacaggcgagttatgccttaggtcaaggttctgtgttcactcagagtgataataattcta 1427  
Db 662 tacaggcgagttatgccttaggtcaaggttctgtgttcactcagagtgataataattcta 721  
Qy 1428 gaccggcctatacaccataacccagcgactagaccaccaagtttcaagccaccaataata 1487  
Db 722 gaccggcctatacaccataacccagcgactagaccaccaagtttcaagccaccaataata 781  
Qy 1488 cgaagcggggtccactttccattggttagtaactctatcaaacgttaaaatcatcaatgtcg 1547  
Db 782 cgaagcggggtccactttccattggttagtaactctatcaaacgttaaaatcatcaatgtcg 841  
Qy 1548 gtgcaggtgttaataaaacagcggtcgcaatgtggcacagctagaagcggtgtggaagt 1607  
Db 842 gtgcaggtgttaataaaacagcggtcgcaatgtggcacagctagaagcggtgtggaagt 901  
Qy 1608 gggctaaggagcgtagaattacttctcaggggtgatataacagctactgacgtaaaaatag 1667  
Db 902 gggctaaggagcgtagaattacttctcaggggtgatataacagctactgacgtaaaaatag 961  
Qy 1668 gtttggataatacttttaacttaataaggtggtgcagagaccaacgcatttaaccgataata 1727  
Db 962 gtttggataatacttttaacttaataaggtggtgcagagaccaacgcatttaaccgataata 1021  
Qy 1728 atatcgggtgtgtaaaagagggctgataataggtgtctgaaagttaaaacttgcataaaactt 1787  
Db 1022 atatcgggtgtgtaaaagagggctgataataggtgtctgaaagttaaaacttgcataaaactt 1081

Qy 1788 taaacaattcttactgaggtgaatacaactacattaaatgccacaaccacagttaaaggtag 1847  
Db 1082 taaacaattcttactgaggtgaatacaactacattaaatgccacaaccacagttaaaggtag 1141  
Qy 1848 gtagtagtagtagtactacagctgaattatigagtatagtttaaacctttaccagagcca 1907  
Db 1142 gtagtagtagtagtactacagctgaattatigagtatagtttaaacctttaccagagcca 1201  
Qy 1908 atacaggcagtcacaagcaaacgcaaacgcgtctatggcggttaattgggtgaagtttacta 1967  
Db 1202 atacaggcagtcacaagcaaacgcaaacgcgtctatggcggttaattgggtgaagtttacta 1261  
Qy 1968 ataatacagaacaaacacagcagcaatcgccactactcgttattaccagagataaaattggct 2027  
Db 1262 ataatacagaacaaacacagcagcaatcgccactactcgttattaccagagataaaattggct 1321  
Qy 2028 ttgctcagataggtgtagtattgataaaaaaacagcaccatttttgataaaaaaacactta 2087  
Db 1322 ttgctcagataggtgtagtattgataaaaaaacagcaccatttttgataaaaaaacactta 1381  
Qy 2088 aagtggggtagtggttgcgaattaccatagacaatggcattgatgcaggttaataaaagatca 2147  
Db 1382 aagtggggtagtggttgcgaattaccatagacaatggcattgatgcaggttaataaaagatca 1441  
Qy 2148 gtaattcttgcacaaggtgtagtgctaacgatgcggttaccatcgcaacagctcaaaagcgg 2207  
Db 1442 gtaattcttgcacaaggtgtagtgctaacgatgcggttaccatcgcaacagctcaaaagcgg 1501  
Qy 2208 ccaagctcacttttaaacgcagcgctggcactcaggtgcacacctactgaaatatcagttg 2267  
Db 1502 ccaagctcacttttaaacgcagcgctggcactcaggtgcacacctactgaaatatcagttg 1561  
Qy 2268 atgttaagtgggcaatgttaccgcccacaacttaacaattggcggtgaaacaccagcagc 2327  
Db 1562 atgttaagtgggcaatgttaccgcccacaacttaacaattggcggtgaaacaccagcagc 1621  
Qy 2328 ttaacagtgatggcactagtgataaaatttagttaaagggttagtggtacgacaataatagct 2387  
Db 1622 ttaacagtgatggcactagtgataaaatttagttaaagggttagtggtacgacaataatagct 1681  
Qy 2388 tagttaccgcgcgaacatttgcgaagctatctaaatgaagttcaatcgaaacgctgacagtg 2447  
Db 1682 tagttaccgcgcgaacatttgcgaagctatctaaatgaagttcaatcgaaacgctgacagtg 1741  
Qy 2448 ctctacaaggtttaccggttaaagaagacgatgtagcgcccaacgctatcaccgtagg 2507  
Db 1742 ctctacaaggtttaccggttaaagaagacgatgtagcgcccaacgctatcaccgtagg 1801  
Qy 2508 ctacaagatcacgacaaaaaaatgccggcgagtcagcatctttaaactcaaaaggttaaaaaag 2567  
Db 1802 ctacaagatcacgacaaaaaaatgccggcgagtcagcatctttaaactcaaaaggttaaaaaag 1861  
Qy 2568 gtctaaacggttgcctaccacaaaaaagatgtagcgtttacccttggcggttagcacaagatagc 2627  
Db 1862 gtctaaacggttgcctaccacaaaaaagatgtagcgtttacccttggcggttagcacaagatagc 1921  
Qy 2628 gtctgacattggcacaacaccccttaaacacagtagggttgaactgttaaaagataccaagc 2687  
Db 1922 gtctgacattggcacaacaccccttaaacacagtagggttgaactgttaaaagataccaagc 1981  
Qy 2688 aacaaatccaagtcggtgtaattggcattaaatttactaaatgtaattgtagtaataccag 2747  
Db 1982 aacaaatccaagtcggtgtaattggcattaaatttactaaatgtaattgtagtaataccag 2041  
Qy 2748 gtactggcattgcaaaatccgctcattaccagagataaaattggcttctgtggttctg 2807  
Db 2042 gtactggcattgcaaaatccgctcattaccagagataaaattggcttctgtggttctg 2101  
Qy 2808 atggtgcagttgatacaaaacaccccttattcttgacaagacaagctacaagttggcagatg 2867  
Db 2102 atggtgcagttgatacaaaacaccccttattcttgacaagacaagctacaagttggcagatg 2161

QY 2868 ttaagattaccacactggcattaaagcaggtggtaagccatcacagggtgtcccaaa 2927  
DB 2162 ttaagattaccacactggcattaaagcaggtggtaagccatcacagggtgtcccaaa 2221  
QY 2928 cactgcctagcattgccgataaagtagccgcaacatagaactgggcaatacaatccaa 2987  
DB 2222 cactgcctagcattgccgataaagtagccgcaacatagaactgggcaatacaatccaa 2281  
QY 2988 acaagacaaaatccacgcgtgcagcattaaatgatataataatcacaggtttacaccta 3047  
DB 2282 acaagacaaaatccacgcgtgcagcattaaatgatataataatcacaggtttacaccta 2341  
QY 3048 aaaaataaacacccattgactttgtctccacttatgaacattgttgaacttgcacatg 3107  
DB 2342 aaaaataaacacccattgactttgtctccacttatgaacattgttgaacttgcacatg 2401  
QY 3108 gcaatgccaccacccgacagtaaacctatgataccgttaacaaaacagtaagtggtat 3167  
DB 2402 gcaatgccaccacccgacagtaaacctatgataccgttaacaaaacagtaagtggtat 2461  
QY 3168 atgattgaaatggatgatacaaacctatcattacacgggactgatgacaataaaaaac 3227  
DB 2462 atgattgaaatggatgatacaaacctatcattacacgggactgatgacaataaaaaac 2521  
QY 3228 ttggcgtcaaaacccacaaactgaacaaaacagtgcttaagtgaatacagcaactaaat 3287  
DB 2522 ttggcgtcaaaacccacaaactgaacaaaacagtgcttaagtgaatacagcaactaaat 2581  
QY 3288 ttaattgtaacttagtgatgaagatgcccctgtttaacgccaagacatgcgcgaataatc 3347  
DB 2582 ttaattgtaacttagtgatgaagatgcccctgtttaacgccaagacatgcgcgaataatc 2641  
QY 3348 taaacccctagcaagaaatcaacacacaaagccagcagcagcagccctcaaaa 3407  
DB 2642 taaacccctagcaagaaatcaacacacaaagccagcagcagcagccctcaaaa 2701  
QY 3408 cctttacgttataaagtagatgaaaataatagctgtagcgcgaacacgcacatcacog 3467  
DB 2702 cctttacgttataaagtagatgaaaataatagctgtagcgcgaacacgcacatcacog 2761  
QY 3468 tgggtcaaaagcaacaaataatcaagtaacacccctaaacacacaaagtgaaacggtc 3527  
DB 2762 tgggtcaaaagcaacaaataatcaagtaacacccctaaacacacaaagtgaaacggtc 2821  
QY 3528 ttaataataaaacgcacaaaatggtacggttaaccttggcattaaacacacaaagcggtc 3587  
DB 2822 ttaataataaaacgcacaaaatggtacggttaaccttggcattaaacacacaaagcggtc 2881  
QY 3588 ttaaagccggcaaaagcacccctaaacacagcggtgctgtctattaaaaacccactggtat 3647  
DB 2882 ttaaagccggcaaaagcacccctaaacacagcggtgctgtctattaaaaacccactggtat 2941  
QY 3648 gcgaacaaatccagtcggtgctgtagcgtggaagtttgcaaggttaataataatggtg 3707  
DB 2942 gcgaacaaatccagtcggtgctgtagcgtggaagtttgcaaggttaataataatggtg 3001  
QY 3708 ttgtaggtgctggtgattgattggcaacactgcattacacagagtgaaattggctttactg 3767  
DB 3002 ttgtaggtgctggtgattgattggcaacactgcattacacagagtgaaattggctttactg 3061  
QY 3768 ggaactaatgctcacttgataaaagcaaacccacccctaaagcaaaagcggcattaacgcag 3827  
DB 3062 ggaactaatgctcacttgataaaagcaaacccacccctaaagcaaaagcggcattaacgcag 3121  
QY 3828 gtggtaaaagattaccacacttaacatcaggtgagattgcccaaaacagccatgatgctg 3887  
DB 3122 gtggtaaaagattaccacacttaacatcaggtgagattgcccaaaacagccatgatgctg 3181  
QY 3888 tgacagggcggaagtttatgatttaaaccggaacttgaaacatacagcagctactg 3947  
DB 3182 tgacagggcggaagtttatgatttaaaccggaacttgaaacatacagcagctactg 3241  
QY 3948 ccaaaacagcacaaaactcattacacgaattctcagtagcagatgaacaaaggtataaact 4007

DB 3242 ccaaaacagcacaaaactcattacacgaattctcagtagcagatgaacaaaggtataaact 3301  
QY 4008 ttacgggttagtaacccttactccagttatgacacactcaaaagacccctgattgtatcact 4067  
DB 3302 ttacgggttagtaacccttactccagttatgacacactcaaaagacccctgattgtatcact 3361  
QY 4068 ttgacgggtgaacacggcattaccaccaaaggtataaataaagggtggtgctggtggcattg 4127  
DB 3362 ttgacgggtgaacacggcattaccaccaaaggtataaataaagggtggtgctggtggcattg 3421  
QY 4128 accaaaccaaaggcttaaccacgocctaaagctgacccgtgggttaataataatggcgaaggca 4187  
DB 3422 accaaaccaaaggcttaaccacgocctaaagctgacccgtgggttaataataatggcgaaggca 3481  
QY 4188 ttgtcattgacagcaaaatggtcaaaataccatcacaggtactgaagcaacactctagcta 4247  
DB 3482 ttgtcattgacagcaaaatggtcaaaataccatcacaggtactgaagcaacactctagcta 3541  
QY 4248 atgtttaccaatgataaaggtagcgtacgcacacacagacagggcaatataatcaaaagacg 4307  
DB 3542 atgtttaccaatgataaaggtagcgtacgcacacacagacagggcaatataatcaaaagacg 3601  
QY 4308 aagcaaaaacccgtgcccgcagcattgtgtatgctgaagcgcagcgttttaacttgcgaag 4367  
DB 3602 aagcaaaaacccgtgcccgcagcattgtgtatgctgaagcgcagcgttttaacttgcgaag 3661  
QY 4368 gcaatgggtgaacgggtgactttgtctccacttatgacacgcgtcaactttgcggtggca 4427  
DB 3662 gcaatgggtgaacgggtgactttgtctccacttatgacacgcgtcaactttgcggtggca 3721  
QY 4428 atgcaacacccgttaaggtgacccatgatacagcattgaagtttaaagataaaaaccca 4547  
DB 3722 atgcaacacccgttaaggtgacccatgatacagcattgaagtttaaagataaaaaccca 3841  
QY 4548 ccacattgcccaggtactggcacaggtgcttaataataatggtcccttaagcaatcaagctactg 4607  
DB 3842 ccacattgcccaggtactggcacaggtgcttaataataatggtcccttaagcaatcaagctactg 3901  
QY 4608 gggatgcgctgttcagggcaggtgatactgctctatcattcaacacacttatcttgcggaca 4667  
DB 3902 gggatgcgctgttcagggcaggtgatactgctctatcattcaacacacttatcttgcggaca 3961  
QY 4668 tccaaactgcgaaggggcaagccaagccaagccaacactcagcaggtatgttggatgctgatg 4727  
DB 3962 tccaaactgcgaaggggcaagccaagccaagccaacactcagcaggtatgttggatgctgatg 4021  
QY 4728 gcaataaggtcattctatgacagtaacgataacagtaactactcaagccaacaaatgagtgga 4787  
DB 4022 gcaataaggtcattctatgacagtaacgataacagtaactactcaagccaacaaatgagtgga 4081  
QY 4788 cagttgataaaaccaaagaagttgccaagacaaactggtgcgcccaagccccacccccag 4847  
DB 4082 cagttgataaaaccaaagaagttgccaagacaaactggtgcgcccaagccccacccccag 4141  
QY 4848 atggcacattggctcacaatgaaatgcaaatcagctcattcaacaaagcaacagtaaatgatg 4907  
DB 4142 atggcacattggctcacaatgaaatgcaaatcagctcattcaacaaagcaacagtaaatgatg 4201  
QY 4908 ccaataaaagcagcaggtcattcaatgaagacaaacgcctttgttaaaggacttgaaaagccg 4967  
DB 4202 ccaataaaagcagcaggtcattcaatgaagacaaacgcctttgttaaaggacttgaaaagccg 4261  
QY 4968 cttctgataacaaaaccaaagcgcagtaactggtgggtgattttaaagtcgcgttgcgc 5027  
DB 4262 cttctgataacaaaaccaaagcgcagtaactggtgggtgattttaaagtcgcgttgcgc 4321  
QY 5028 aaacacgcctgaccttggcaggggatacagcacaacgcgtctaaaaacttggcgagactt 5087

Dbb 4322 aaacacccgctgacaccttgcaggggatacagggcacaacggtctaaacacccgctggcgagactt 4381  
Qy 5088 tgacctcaaaagtgggcaaacagacaccaataagctaaacggataataaacatcgtgtgg 5147  
Dbb 4382 tgacctcaaaagtgggcaaacagacaccaataagctaaacggataataaacatcgtgtgg 4441  
Qy 5148 tagcaggtactgtaggttcaactgtcaaaacttgcaaaagacactcaacaaatttaaacagc 5207  
Dbb 4442 tagcaggtactgtaggttcaactgtcaaaacttgcaaaagacactcaacaaatttaaacagc 4501  
Qy 5208 ttaatgcaggtggcaccacaaattgatgacaaagcggtgtctttttagactcaagcggtc 5267  
Dbb 4502 ttaatgcaggtggcaccacaaattgatgacaaagcggtgtctttttagactcaagcggtc 4561  
Qy 5268 aagccaaagcaaacacccctgtgtaagtgcgaatggctggactgggtggcgaaggtca 5327  
Dbb 4562 aagccaaagcaaacacccctgtgtaagtgcgaatggctggactgggtggcgaaggtca 4621  
Qy 5328 tcagtaattgtggcacaaggcacaagataccgacgctgccaatgtcaacacagttaaacg 5387  
Dbb 4622 tcagtaattgtggcacaaggcacaagataccgacgctgccaatgtcaacacagttaaacg 4681  
Qy 5388 aagtcacgcaactgtgtggctgtgtaagtctgtaataagataacgctgacgccaatcagg 5447  
Dbb 4682 aagtcacgcaactgtgtggctgtgtaagtctgtaataagataacgctgacgccaatcagg 4741  
Qy 5448 taacattgcgcacatcaaaaagaccacaaattcaggttcatcatctaaacgcgactgtca 5507  
Dbb 4742 taacattgcgcacatcaaaaagaccacaaattcaggttcatcatctaaacgcgactgtca 4801  
Qy 5508 tcaagcaggcagcgttacttggcgttaaggtataacgataccgcaaaaaacttgccactg 5567  
Dbb 4802 tcaagcaggcagcgttacttggcgttaaggtataacgataccgcaaaaaacttgccactg 4861  
Qy 5568 gtgtatcacagtggtggtgtgataaaacgcaacgctaaacgagcttaagcaaatgttt 5627  
Dbb 4862 gtgtgtatacaagtggtggtgtgataaaacgcaacgctaaacgagcttaagcaaatgttt 4921  
Qy 5628 ggggtcaaaaccccaaaagatgagcagcaaaaagccctgctgcacacttataacgccgag 5687  
Dbb 4922 ggggtcaaaaccccaaaagatgagcagcaaaaagccctgctgcacacttataacgccgag 4981  
Qy 5688 gtcagaccaaactatttgacacaaaccccgagaaagccattgacagaataaatgaacaag 5747  
Dbb 4982 gtcagaccaaactatttgacacaaaccccgagaaagccattgacagaataaatgaacaag 5041  
Qy 5748 gtatccgctcttccatgtcaacatggcaatcaagcctgtgtgtacaaagcgctaaacg 5807  
Dbb 5042 gtaaccgctcttccatgtcaacatggcaatcaagcctgtgtgtacaaagcgctaaacg 5101  
Qy 5808 gcatgtactcaagtgcctcagggcaagcactcagtggtggcagtaggtttccagggccaagcgag 5867  
Dbb 5102 gcatgtactcaagtgcctcagggcaagcactcagtggtggcagtaggtttccagggccaagcgag 5161  
Qy 5868 atggtgaagcccgctgtgccatagggagacaaaaccccaagcggaacccaatcccatcgcca 5927  
Dbb 5162 atggtgaagcccgctgtgccatagggagacaaaaccccaagcggaacccaatcccatcgcca 5221  
Qy 5928 tcggtgataacgcacaagccagcggtcatcaatccatcgccatcggtacagggcaatgtgg 5987  
Dbb 5222 tcggtgataacgcacaagccagcggtcatcaatccatcgccatcggtacagggcaatgtgg 5281  
Qy 5988 tagcaggttaagcactctggtgccatcggcgacccaagcactgttaagcgtgataacagtt 6047  
Dbb 5282 tagcaggttaagcactctggtgccatcggcgacccaagcactgttaagcgtgataacagtt 5341  
Qy 6048 acagtggtggtaataacacacagtttaaccgattgcacactcaaacggatgtctttggtgg 6107  
Dbb 5342 acagtggtggtaataacacacagtttaaccgattgcacactcaaacggatgtctttggtgg 5401  
Qy 6108 gcaataacatcacctgacccaagaaagtaactcgttgccttaggttcaaacctcgtccatca 6167  
Dbb 5402 gcaataacatcacctgacccaagaaagtaactcgttgccttaggttcaaacctcgtccatca 5461

Qy 6168 gtgcaggcacacgcagcgacacaaagcccaaaaaactctgcagggcacagcaggtacaacca 6227  
Dbb 5462 gtgcaggcacacgcagcgacacaaagcccaaaaaactctgcagggcacagcaggtacaacca 5521  
Qy 6228 ccacagcaggtgcacaacggtacggttaaagccttgcctggacaaacgcggtgtggtggc 6287  
Dbb 5522 ccacagcaggtgcacaacggtacggttaaagccttgcctggacaaacgcggtgtggtggc 5581  
Qy 6288 tctccgtgggtgcctcaggtgctgaacccgtatccaaaatgtgcagcaggtgaggtca 6347  
Dbb 5582 tctccgtgggtgcctcaggtgctgaacccgtatccaaaatgtgcagcaggtgaggtca 5641  
Qy 6348 gtgcaccagcacccgatcggttcaatgtgaccagttgtacaaaagccaccccaaacgattg 6407  
Dbb 5642 gtgcaccagcacccgatcggttcaatgtgaccagttgtacaaaagccaccccaaacgattg 5701  
Qy 6408 ccaacgcaacaaatgagcttgaccatcgtatccaccacaaagcaaaaatgaaggccaatgag 6467  
Dbb 5702 ccaacgcaacaaatgagcttgaccatcgtatccaccacaaagcaaaaatgaaggccaatgag 5761  
Qy 6468 ggatttcacagcagtgatggcggtccatgcacagcctcacattcctgcagatcca 6527  
Dbb 5762 ggatttcacagcagtgatggcggtccatgcacagcctcacattcctgcagatcca 5821  
Qy 6528 tggttaccgggggtatttgcaccccaaacggtcaaggtgcggtggcagtggtgactgtcga 6587  
Dbb 5822 tggttaccgggggtatttgcaccccaaacggtcaaggtgcggtggcagtggtgactgtcga 5881  
Qy 6588 agctgcggataatgtgtaaatgggtatttaaaatcaatgggttcagccgatacccaagcc 6647  
Dbb 5882 agctgcggataatgtgtaaatgggtatttaaaatcaatgggttcagccgatacccaagcc 5941  
Qy 6648 atgtagggcgcgagtggtgcaggtttcacttttaagccataaatcgcaagattttac 6707  
Dbb 5942 atgtagggcgcgagtggtgcaggtttcacttttaagccataaatcgcaagattttac 6001  
Qy 6708 ttaaaaaatcaatctcaccatagttgtataaaaacagcagcagcagtcattactgat 6767  
Dbb 6002 ttaaaaaatcaatctcaccatagttgtataaaaacagcagcagcagtcattactgat 6061  
Qy 6768 gctgagtgtttttatcaacttaaacattttaccgctcaagtgattcttttaccatgac 6827  
Dbb 6062 gctgagtgtttttatcaacttaaacattttaccgctcaagtgattcttttaccatgac 6121  
Qy 6828 caaatcgccattgacataggttaaaacttatigagtaaattttatcaatgtattgttaga 6887  
Dbb 6122 caaatcgccattgacataggttaaaacttatigagtaaattttatcaatgtattgttaga 6181  
Qy 6888 tatggttaaaaattgtgccattgacccaaaaatgaccgattttatccgaaaaattttctgatt 6947  
Dbb 6182 tatggttaaaaattgtgccattgacccaaaaatgaccgattttatccgaaaaattttctgatt 6241  
Qy 6948 atgacccgttgacctgca 6965  
Dbb 6242 atgacccgttgacctgca 6259

## RESULT 5

AAF59103  
ID AAF59103 standard; DNA; 6144 BP.

.xx  
AC AAF59103;

XX  
DT 24-APR-2001 (first entry)

XX  
DE M. catarrhalis strain 4223 genomic 200kDa coding sequence SEQ ID NO:6.

XX  
KW Moraxella catarrhalis strain 4223; major outer membrane protein;

XX  
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

XX  
KW otitis media; detection; ds.

OS Moraxella catarrhalis.



Db 1501 gatgtgatgaaacaaagcaccacatatttgataaaaaacaaactaaagtgggtagtgctt 1560  
Qy 2102 gcaattaccatagacaatggcattgatgcaggtaataaaaaagatcagtaattcttgcacaa 2161  
Db 1561 gcaattaccatagacaatggcattgatgcaggtaataaaaaagatcagtaattcttgcacaa 1620  
Qy 2162 ggtagcagtgtaacagatgcgggttacatcgaaacagctcaaaagccgaagccctacttta 2221  
Db 1621 ggtagcagtgtaacagatgcgggttacatcgaaacagctcaaaagccgaagccctacttta 1680  
Qy 2222 aacgcagcgtggcatcagtgatcacaccctactgaaatatatcagttgattgcttaagatggc 2281  
Db 1681 aacgcagcgtggcatcagtgatcacaccctactgaaatatatcagttgattgcttaagatggc 1740  
Qy 2282 aatgtaccgcccacacttaacattggcgtgaaacccacggagcttaacagtgatggc 2341  
Db 1741 aatgtaccgcccacacttaacattggcgtgaaacccacggagcttaacagtgatggc 1800  
Qy 2342 actagtataaatttagttagggtagtggtagcaaacataagcttagttaccgccgaa 2401  
Db 1801 actagtataaatttagttagggtagtggtagcaaacataagcttagttaccgccgaa 1860  
Qy 2402 catttgcagactatcaaatgaatgaatcaatcgaaacggtgacagtgtctacaaagcttt 2461  
Db 1861 catttgcagactatcaaatgaatgaatcaatcgaaacggtgacagtgtctacaaagcttt 1920  
Qy 2462 accgttaaaagaagacgatgtagcgccaaacgctatcacctgggttaagaatcagaca 2521  
Db 1921 accgttaaaagaagacgatgtagcgccaaacgctatcacctgggttaagaatcagaca 1980  
Qy 2522 aaaaatccggcgagtcagcatcttaaaactcaaaaggttaaaaacggtcttaacogtttgc 2581  
Db 1981 aaaaatccggcgagtcagcatcttaaaactcaaaaggttaaaaacggtcttaacogtttgc 2040  
Qy 2582 accaaaaaagatgtagcgtttaccttggccttagcgaagatagcgtctgaccattggc 2641  
Db 2041 accaaaaaagatgtagcgtttaccttggccttagcgaagatagcgtctgaccattggc 2100  
Qy 2642 aaaaagcccttaaaacacgatggctgactgtttaagataccaaacgaacaaatccaagtc 2701  
Db 2101 aaaaagcccttaaaacacgatggctgactgtttaagataccaaacgaacaaatccaagtc 2160  
Qy 2702 ggtcctaagtgcattaaatttacttaattgaatggtagtaataccagtlactggcatggca 2761  
Db 2161 ggtcctaagtgcattaaatttacttaattgaatggtagtaataccagtlactggcatggca 2220  
Qy 2762 aataccgctgcattaccagagataaaattggcttgcgtgtctgattggtgcagttgat 2821  
Db 2221 aataccgctgcattaccagagataaaattggcttgcgtgtctgattggtgcagttgat 2280  
Qy 2822 acaacaaaccttatcttgatcaagacaagctacaagtggcaattggcaattgtaagattaccaac 2881  
Db 2281 acaacaaaccttatcttgatcaagacaagctacaagtggcaattgtaagattaccaac 2340  
Qy 2882 actggcattaacgaggtggtaaaagccatcacagggctgtcccaacacactgccttagatt 2941  
Db 2341 actggcattaacgaggtggtaaaagccatcacagggctgtcccaacacactgccttagatt 2400  
Qy 2942 gccgatcaaatgtagcccaacatagaactggcgaatacaatccaagacaagacaataccc 3001  
Db 2401 gccgatcaaatgtagcccaacatagaactggcgaatacaatccaagacaagacaataccc 2460  
Qy 3002 aacgctgccagcattatgatatataatacagcgttttaaccttaaaaaataatacaaac 3061  
Db 2461 aacgctgccagcattatgatatataatacagcgttttaaccttaaaaaataatacaaac 2520  
Qy 3062 cccattgactttgtctccacttatgacattgttgactttgccaatggcaatggccaccacc 3121  
Db 2521 cccattgactttgtctccacttatgacattgttgactttgccaatggcaatggccaccacc 2580  
Qy 3122 gccacagtaaccatgataccgcttaacaaaacagtaaaagtggtatatgtgtaattg 3181  
Db 2581 gccacagtaaccatgataccgcttaacaaaacagtaaaagtggtatatgtgtaattg 2640

Qy 3182 gatgatacaaccattcatcttaacaggcactgatgacaataaaaaacttggcgtcaaaaacc 3241  
Db 2641 gatgatacaaccattcatcttaacaggcactgatgacaataaaaaacttggcgtcaaaaacc 2700  
Qy 3242 accaaactgaacaaaaaaagtgttaattggtaatacagcaactaacttttaattgtaactct 3301  
Db 2701 accaaactgaacaaaaaaagtgttaattggtaatacagcaactaacttttaattgtaactct 2760  
Qy 3302 agtgatgaagatgccttgttaacgcaaaagacatgcgcgaaatcttaaacacacctagcc 3361  
Db 2761 agtgatgaagatgccttgttaacgcaaaagacatgcgcgaaatcttaaacacacctagcc 2820  
Qy 3362 aaggaattcacacacccaaagggcacagacacccgcttacaacactttaccggttaaa 3421  
Db 2821 aaggaattcacacacccaaagggcacagacacccgcttacaacactttaccggttaaa 2880  
Qy 3422 aaggtagatgaaataataatgtgatgaccccaacgcatcacccgtgggttcaaaaagac 3481  
Db 2881 aaggtagatgaaataataatgtgatgaccccaacgcatcacccgtgggttcaaaaagac 2940  
Qy 3482 gcaataatcaagtcaaacacccctaacactcaaggtgaaacggtcttaattatataaac 3541  
Db 2941 gcaataatcaagtcaaacacccctaacactcaaggtgaaacggtcttaattatataaac 3000  
Qy 3542 gacaaaaatgtagcgtttaccttggcattaacaccacaagcgtctttaaagccggcaaa 3601  
Db 3001 gacaaaaatgtagcgtttaccttggcattaacaccacaagcgtctttaaagccggcaaa 3060  
Qy 3602 agcccttaaacgacggttgcttctattaaaaacccactggtagcgacaacataccaa 3661  
Db 3061 agcccttaaacgacggttgcttctattaaaaacccactggtagcgacaacataccaa 3120  
Qy 3662 gtcggtgctgtagcgtgaaagtgttgcgaaggttaataataatggtgtgtgtaggtgctggc 3721  
Db 3121 gtcggtgctgtagcgtgaaagtgttgcgaaggttaataataatggtgtgtgtaggtgctggc 3180  
Qy 3722 attgatggcaacactgcattaccagagatgaaattggtcttactgggactaaatggctca 3781  
Db 3181 attgatggcaacactgcattaccagagatgaaattggtcttactgggactaaatggctca 3240  
Qy 3782 ctgtataaaaaacacccaccctaaagcaaaagacggtcattaaacgaggtggtaaaaagatt 3841  
Db 3241 ctgtataaaaaacacccaccctaaagcaaaagacggtcattaaacgaggtggtaaaaagatt 3300  
Qy 3842 accaaacttaacatcaggtgagattgcccacaaacagccatgctgtgacagggcggaag 3901  
Db 3301 accaaacttaacatcaggtgagattgcccacaaacagccatgctgtgacagggcggaag 3360  
Qy 3902 atttatgatttaaaaacccgaacttgaacacaaaatcagcagtactgccaaaacacagcaca 3961  
Db 3361 atttatgatttaaaaacccgaacttgaacacaaaatcagcagtactgccaaaacacagcaca 3420  
Qy 3962 aactcattacacgaattctcagtagcagatgaacaaaggttaataactttacaggttagtaac 4021  
Db 3421 aactcattacacgaattctcagtagcagatgaacaaaggttaataactttacaggttagtaac 3480  
Qy 4022 ccttactccagttagacacctcaaaacccctctgattgcatcacctttgacggtgaaaac 4081  
Db 3481 ccttactccagttagacacctcaaaacccctctgattgcatcacctttgacggtgaaaac 3540  
Qy 4082 ggcattaccacacaggttaataaaaggtgtgtggtgtggcatttgacccaacccaagggc 4141  
Db 3541 ggcattaccacacaggttaataaaaggtgtgtggtgtggcatttgacccaacccaagggc 3600  
Qy 4142 ttaaccacgctaaagctgaacggtgggttaataataatggcaaaaggttgcattgacagc 4201  
Db 3601 ttaaccacgctaaagctgaacggtgggttaataataatggcaaaaggttgcattgacagc 3660  
Qy 4202 caaatggtcaaaatcaccatcacaggaactaacactctagctaatgtttaccaatgat 4261  
Db 3661 caaatggtcaaaatcaccatcacaggaactaacactctagctaatgtttaccaatgat 3720

QY 4262 aaaggtagcgtacgcacacagacagggccaataataatcaaaagcgaagacacaaacccgt 4321  
|||||  
Db 3721 aaaggtagcgtacgcacacagacagggccaataataatcaaaagcgaagacacaaacccgt 3780  
QY 4322 gccgcagcattgttgatgtcgaagcagcaggtttaaacttgaagcgaatgggtgaagcg 4381  
|||||  
Db 3781 gccgcagcattgttgatgtcgaagcagcaggtttaaacttgaagggcaatgggtgaagcg 3840  
QY 4382 gttgactttgtcccaacttatgacacccgttcaactttgcgattggaatggccaccacgct 4441  
|||||  
Db 3841 gttgactttgtcccaacttatgacacccgttcaactttgcgattggaatggccaccacgct 3900  
QY 4442 aaagtgacctatgatgacacagcgaacaaacccagtaaaagtgtgtctatgatgtcaattgcat 4501  
|||||  
Db 3901 aaagtgacctatgatgacacagcgaacaaacccagtaaaagtgtgtctatgagtcaattgcat 3960  
QY 4502 gatcaaacattgaagttaaagataaaaaacttggcgttaaaacaccaccacattgaccagt 4561  
|||||  
Db 3961 gatcaaacattgaagttaaagataaaaaacttggcgttaaaacaccaccacattgaccagt 4020  
QY 4562 actggcacaggtgctaataattggccctaagcaatacaagcttacttgcgagcgtgtgttc 4621  
|||||  
Db 4021 actggcacaggtgctaataattggccctaagcaatacaagcttacttgcgagcgtgtgttc 4080  
QY 4622 aaggcagtgatcgttgcctatcaaacaccttatctggcgacatccaaactgcacaa 4681  
|||||  
Db 4081 aaggcagtgatcgttgcctatcaaacaccttatctggcgacatccaaactgcacaa 4140  
QY 4682 gggcgaagcgaacgaacactcagcaggcta tgtggatgctgatggcgaataaaggtcacc 4741  
|||||  
Db 4141 gggcgaagcgaacgaacactcagcaggcta tgtggatgctgatggcgaataaaggtcacc 4200  
QY 4742 tatgacagtcaccgatacaaaagtactatcaagccaaataatgatggcacagtgtgataaaacc 4801  
|||||  
Db 4201 tatgacagtcaccgatacaaaagtactatcaagccaaataatgatggcacagtgtgataaaacc 4260  
QY 4802 aaagaagtgcgaagacaaacttgtgcgccaaagcccaaacccacagatggcacattggct 4861  
|||||  
Db 4261 aaagaagtgcgaagacaaacttgtgcgccaaagcccaaacccacagatggcacattggct 4320  
QY 4862 caaatgaatgtcaaatcagtcattaaacaagaacaaagtaaatatgccataaaagcga 4921  
|||||  
Db 4321 caaatgaatgtcaaatcagtcattaaacaagaacaaagtaaatatgccataaaagcga 4380  
QY 4922 ggcataatgaagacaaacgcttctgttaaaggacttgaaagacccgcttctgtatacaaa 4981  
|||||  
Db 4381 ggcataatgaagacaaacgcttctgttaaaggacttgaaagacccgcttctgtatacaaa 4440  
QY 4982 accaaaaacgcgcagtaacttgggtgatttaaatccgttgcacaaacacccgctgacc 5041  
|||||  
Db 4441 accaaaaacgcgcagtaacttgggtgatttaaatccgttgcacaaacacccgctgacc 4500  
QY 5042 ttctcaggggatacagggcacacagcgttaaaaaactggcgagactttgacacataaaggt 5101  
|||||  
Db 4501 ttctcaggggatacagggcacacagcgttaaaaaactggcgagactttgacacataaaggt 4560  
QY 5102 gggcacaacagacaccaataaagtaaccgataataacatccgttggtagcaggtactgat 5161  
|||||  
Db 4561 gggcacaacagacaccaataaagtaaccgataataacatccgttggtagcaggtactgat 4620  
QY 5162 ggtctcactgtcaaaacttgcaaaagacacctaaccaatcttaacagcgttaattgcaggtggc 5221  
|||||  
Db 4621 ggtctcactgtcaaaacttgcaaaagacacctaaccaatcttaacagcgttaattgcaggtggc 4680  
QY 5222 accaaaaattgatgacaaagggcgtgtctttttagactcaagcggltcaagccaaagcgaac 5281  
|||||  
Db 4681 accaaaaattgatgacaaagggcgtgtctttttagactcaagcggltcaagccaaagcgaac 4740  
QY 5282 accctgtactgaagtcacaaatgggctggacccctgggtggcgaaggtcatcagtaattgggc 5341  
|||||  
Db 4741 accctgtactgaagtcacaaatgggctggacccctgggtggcgaaggtcatcagtaattgggc 4800  
QY 5342 aaaggcacaaaaagataccgcagcgtgccaatgtacaacaggttaaacgaagtcagcaacttg 5401

Db 4801 aaaggcacaaaaagataccgcagcgtgccaattgataacacagttataacgaagtcagcaacttg 4860  
QY 5402 ttgggtcttgtaataactggttaataacgctgacgcgaataacacaggtataacatttgcgcgac 5461  
|||||  
Db 4861 ttgggtcttgtaataactggttaataacgctgacgcgaataacacaggtataacatttgcgcgac 4920  
QY 5462 atcaaaaaagaccccaattcaggtttcatctatctataacccgcacgtgtcatcaaaagcagcag 5521  
|||||  
Db 4921 atcaaaaaagaccccaattcaggtttcatctatctataacccgcacgtgtcatcaaaagcagcag 4980  
QY 5522 gtacttggcgttaaaagtaataacgataacgaaaaaacttgcacactggtgtgtatcacaaagtg 5581  
|||||  
Db 4981 gtacttggcgttaaaagtaataacgataacgaaaaaacttgcacactggtgtgtgtatcacaaagtg 5040  
QY 5582 ggcgtgataaaaagcgcgaacgcgttaacgcggaatttaagcaatttttgggttcaaaaccccaa 5641  
|||||  
Db 5041 ggcgtgataaaaagcgcgaacgcgttaacgcggaatttaagcaatttttgggttcaaaaccccaa 5100  
QY 5642 aaagatggcagcaaaaaagccctgctgcacattataacccgcaggtcagacacactat 5701  
|||||  
Db 5101 aaagatggcagcaaaaaagccctgctgcacattataacccgcaggtcagacacactat 5160  
QY 5702 ttgaccaaacacccgcgacgaagccattgacagaataaatagaacaaaggtatccgctttctc 5761  
|||||  
Db 5161 ttgaccaaacacccgcgacgaagccattgacagaataaatagaacaaaggtatccgctttctc 5220  
QY 5762 catgtcaacgataagcaatcaaaagccctgctgcacattataacccgcaggtcagacacactat 5821  
|||||  
Db 5221 catgtcaacgataagcaatcaaaagccctgctgcacattataacccgcaggtcagacacactat 5280  
QY 5822 gcctcaggaacagcactcagtgccagataggtttccaggccaagggcagatgggtgaagccgccc 5881  
|||||  
Db 5281 gcctcaggaacagcactcagtgccagataggtttccaggccaagggcagatgggtgaagccgccc 5340  
QY 5882 gttgcataaggcagacaaaccccaagcaggaaccccaatccatccatcgtgcatacgttgataacgca 5941  
|||||  
Db 5341 gttgcataaggcagacaaaccccaagcaggaaccccaatccatccatcgttgataacgca 5400  
QY 5942 caagccacggcgagatcaatccatccatccgtgtaacaggaataatggtagcaggttaagcac 6001  
|||||  
Db 5401 caagccacggcgagatcaatccatccatccgtgtaacaggaataatggtagcaggttaagcac 5460  
QY 6002 tctgtgcatcggcgaccccaagcactgttgaaggtctgataacagttacagttgtgggttaat 6061  
|||||  
Db 5461 tctgtgcatcggcgaccccaagcactgttgaaggtctgataacagttacagttgtgggttaat 5520  
QY 6062 aacaaacagtttaccgagtcgcaactcaaacccgctgtcttgggtgggcaataaatacacc 6121  
|||||  
Db 5521 aacaaacagtttaccgagtcgcaactcaaacccgctgtcttgggtgggcaataaatacacc 5580  
QY 6122 gtgacccgaagtaactcgggttgccttaggttcaaaactctgccatcagtcagggcacacac 6181  
|||||  
Db 5581 gtgacccgaagtaactcgggttgccttaggttcaaaactctgccatcagtcagggcacacac 5640  
QY 6182 gaaggcacacaaagcccaaaaaactctgacggcagcaggtataacacccacacagcaggtgca 6241  
|||||  
Db 5641 gcaggcacacaaagcccaaaaaactctgacggcagcaggtataacacccacacagcaggtgca 5700  
QY 6242 accggttacggttaaaggccttgcctggacaaaacggcgttgggtggcgtctccgtgggtgccc 6301  
|||||  
Db 5701 accggttacggttaaaggccttgcctggacaaaacggcgttgggtggcgtctccgtgggtgccc 5760  
QY 6302 tcagggtgtaacgcgcgtatcccaaaatgtggcagcaggttgaggtcagttcagtcaccacagcacc 6361  
|||||  
Db 5761 tcagggtgtaacgcgcgtatcccaaaatgtggcagcaggttgaggtcagttcagtcaccacagcacc 5820  
QY 6362 gatcgggtcaatggttagcaggttgcatacaaaagccaccccaagcatttgcacacgaacaaat 6421  
|||||  
Db 5821 gatcgggtcaatggttagcaggttgcatacaaaagccaccccaagcatttgcacacgaacaaat 5880  
QY 6422 gagcttgaccatcgtatccaccccaaaagcaaaataaaggccaatgcagggatttcatcagcg 6481  
|||||







QY	1727	aatacggctggtgtaaaagagcgtgataaatagtggcttgaaagttaaacttgcataaact	1786
DB	1021	aatacggctggtgtaaaagggctgataacagcggctctgaaagttaaacttgcataaact	1080
QY	1787	ttaaaaactcttactcgagtggaatacaactacattaaatgccacacacacaggttaaagta	1846
DB	1081	ttaaaaactcttactcgagtggaatacaactacattaaatgccacacacacaggttaaagta	1140
QY	1847	ggtagtagtagtagtactacagctgaatattatgtagtgatgtttaacctttaccagccc	1906
DB	1141	ggtagtagtagtagtactacagctgaatattatgtagtgatgtttaacctttaccagccc	1200
QY	1907	aatacaggcagtcctaaagcacaagcaaaccgctctatggcggttaatgggggtgaagttact	1966
DB	1201	aatacaggcagtcctaaagcacaagcaaaccgctctatggcggttaatgggggtgaagttact	1260
QY	1967	ataatgcagaatacaaacagcagcaaatcgggaactactcgtattaccagagataaaatggc	2026
DB	1261	ataatgcagaatacaaacagcagcaaatcgggaactactcgtattaccagagataaaatggc	1320
QY	2027	tttgcctcgagatgggtgattgttgataaaaaaacagcaccattatttggataaaaaaacatt	2086
DB	1321	tttgcctcgagatgggtgattgttgataaaaaaacagcaccattatttggataaaaaaacatt	1380
QY	2087	aaagtgggtagtggtgcgaattaccatagacaaatgggaatgtagtcaggttaataaaagatc	2146
DB	1381	aaagtgggtagtggtgcgaattaccatagacaaatgggaatgtagtcaggttaataaaagatc	1440
QY	2147	agtaactctgcgaaggttagcagtgctaaagatgctgggttaccatcgaaacagctcaagcc	2206
DB	1441	agtaactctgcgaaggttagcagtgctaaagatgctgggttaccatcgaaacagctcaagcc	1500
QY	2207	gccaaagcctactttaaaacgagggcctgggaactcaagtgtcacaactactgaatatcagtt	2266
DB	1501	gccaaagcctactttaaaacgagggcctgggaactcaagtgtcacaactactgaatatcagtt	1560
QY	2267	gatctaaagatgggcaatgttaccgcccacaacttacaacattggcgtgaaaaaccacggag	2326
DB	1561	gatctaaagatgggcaatgttaccgcccacaacttacaacattggcgtgaaaaaccacggag	1620
QY	2327	cttaacagtgatggcactagtataaaatttagtgttaagggtagtggtgcagaacaatagc	2386
DB	1621	cttaacagtgatggcactagtataaaatttagtgttaagggtagtggtgcagaacaatagc	1680
QY	2387	ttagtaccgcggaacatttggcgaagctatctaaatgaagtcaatcgaaacgctgacagt	2446
DB	1681	ttagtaccgcggaacatttggcgaagctatctaaatgaagtcaatcgaaacgctgacagt	1740
QY	2447	gctctacaagctttaccgtttaagaagaagacgatgatgacgccaacgctatcaacgtg	2506
DB	1741	gctctacaagctttaccgtttaagaagaagacgatgatgacgccaacgctatcaacgtg	1800
QY	2507	gctaagaatcacgcaaaaaatccggcgagtcagcatctttaaactccaaggttaaaaaac	2566
DB	1801	gctaagaatcacgcaaaaaatccggcgagtcagcatctttaaactccaaggttaaaaaac	1860
QY	2567	ggcttaacggttgctacccaataaagaatggtagtcggttaactttgggttagccaaagatagc	2626
DB	1861	ggcttaacggttgctacccaataaagaatggtagtcggttaactttgggttagccaaagatagc	1920
QY	2627	ggctctgacctggcaaaagcacccttaaacacagatggctctgactgttaaagataaccaac	2686
DB	1921	ggctctgacctggcaaaagcacccttaaacacagatggctctgactgttaaagataaccaac	1980
QY	2687	gaacaaatccaagtcggtgtaatggcattaaattactaatgtgaatggtagtaatacca	2746
DB	1981	gaacaaatccaagtcggtgtaatggcattaaattactaatgtgaatggtagtaatacca	2040
QY	2747	ggtagtcgcatgtaataacgctcgattaccagagataaaaattggcttggctgggtctct	2806
DB	2041	ggtagtcgcatgtaataacgctcgattaccagagataaaaattggcttggctgggtctct	2100
QY	2807	gatggtcgagttgatacaaaaacaccttatcttgatacgaacgaagctacaagttggcgaat	2866

Db	2101	 gatggtcagttgatcaacaaacctatcttgcatacgaagctacaagttggcaat	2160
Qy	2867	gtaagattaccacactggcattaacgcaggtggtgaagccatcaacagggctgtcccca 	2926
Db	2161	gtaagattacaacactggcattaacgcaggtggtgaagccatcaacagggctgtcccca 	2220
Qy	2927	acactgcctgcatctgcgatcaaaatagccgcagcaacatagaacitgggcaatacaactcaa 	2986
Db	2221	acactgcctgcatctgcgatcaaaatagccgcagcaacatagaacitgggcaatacaactcaa 	2280
Qy	2987	gacaaacaaaatccaacgctgcagcattaatgatataattaaatcacagggctttaacctta 	3046
Db	2281	gacaaagacaaaatccaacgctgcagcattaatgatataattaaatcacagggctttaacctta 	2340
Qy	3047	aaaaataataaacaccccatctgactttgtctccacttatgacattgttgacattgttgacattggccaat 	3106
Db	2341	aaaaataataaacaccccatctgactttgtctccacttatgacattgttgacattgttgacattggccaat 	2400
Qy	3107	ggcaatgccaccacccgcacagtaaaccccatgatccgcttaacaaaaccagtaaaagtggta 	3166
Db	2401	ggcaatgccaccacccgcacagtaaaccccatgatccgcttaacaaaaccagtaaaagtggta 	2460
Qy	3167	tatgatgtgaatgttgatgatacaacattcatcttaacaggcactgtatgacaaataaaaaa 	3226
Db	2461	tatgatgtgaatgttgatgatacaacattcatcttaacaggcactgtatgacaaataaaaaa 	2520
Qy	3227	cttggcgtcaaaaaacacaaactgaacaaaaacaagtctcaatggttaatacacagcaactaac 	3286
Db	2521	cttggcgtcaaaaaacacaaactgaacaaaaacaagtctcaatggttaatacacagcaactaac 	2580
Qy	3287	tttaatgttaactctagtgtatgaagaatgccttgtttaacgccaaagacatcgcgcgaataat 	3346
Db	2581	tttaatgttaactctagtgtatgaagaatgccttgtttaacgccaaagacatcgcgcgaataat 	2640
Qy	3347	ctaaacacccctagccaaggaaattcacaccacaaaggccacgacgacagccacctacaa 	3406
Db	2641	ctaaacacccctagccaaggaaattcacaccacaaaggccacgacgacagccacctacaa 	2700
Qy	3407	acctttaccgttaaaaaggttagatgaataataatgctgtacgaccccaacgcccataccc 	3466
Db	2701	acctttaccgtttaaaaaggttagatgaataataatgctgtacgaccccaacgcccataccc 	2760
Qy	3467	gtgggtcaaaagaaacgcaaaataataagtcacaacccctaacactcaaaaggtgaataacggt 	3526
Db	2761	gtgggtcaaaagaaacgcaaaataataagtcacaacccctaacactcaaaaggtgaataacggt 	2820
Qy	3527	cttaattattaaacccgcaaaaaatggtagctttaccctttggcatttaacaccacaagcgggt 	3584
Db	2821	cttaattattaaacccgcaaaaaatggtagctttaccctttggcatttaacaccacaagcgggt 	2880
Qy	3587	cttaaacgcgcgcaaaagcacccctaaacgaagctggctgtgtctattaaaaacccccactggt 	3644
Db	2881	cttaaacgcgcgcaaaagcacccctaaacgaagctggctgtgtctattaaaaacccccactggt 	2940
Qy	3647	agcgaacaaatccaagtgcggtcgtgatggcgtgaagtttgcacaggtttaataataatggt 	3700
Db	2941	agcgaacaaatccaagtgcggtcgtgatggcgtgaagtttgcacaggtttaataataatggt 	3000
Qy	3707	gttgtaggtgctggcattgtagggcacaaactcgcattaccagagatgaaattggcctttact 	3766
Db	3001	gttgtaggtgctggcattgtagggcacaaactcgcattaccagagatgaaattggcctttact 	3060
Qy	3767	gggactaaatggctcatctgtataaaagcaaaacccccactaaagcaaaagcgcattaaacgca 	382
Db	3061	gggactaaatggctcatctgtataaaagcaaaacccccactaaagcaaaagcgcattaaacgca 	312
Qy	3827	ggtagtaaaagatataccaaactcaatcaggtagatgtgccaaaaacagccatgatgct 	388
Db	3121	ggtagtaaaagatataccaaactcaatcaggtagatgtgccaaaaacagccatgatgct 	318
Qy	3887	gtgacagggcggaagatttatgatttaaaaacccgaacttgaaaaacaaaatacagcagtaact 	394

Db 3181 gtcacaggcgcgaagatttatgatttaaaacccgaacttgaaaaacaaatcagcagttact 3240  
Qy 3947 gccaaaacagcacaaaactcattacagaaattctcagtagcagatgacagaggtaataac 4006  
Db 3241 gccaaaacagcacaaaactcattacagaaattctcagtagcagatgacagaggtaataac 3300  
Qy 4007 ttacaggttagtaaccttactccagttatgacacctcacaagacctctgatgtcateacc 4066  
Db 3301 ttacaggttagtaaccttactccagttatgacacctcacaagacctctgatgtcateacc 3360  
Qy 4067 ttgcaggttaaaacgcattaccaccaaagtgaaataaaaggtgtggtgcgtgtggcatt 4126  
Db 3361 ttgcaggttaaaacgcattaccaccaaagtgaaataaaaggtgtggtgcgtgtggcatt 3420  
Qy 4127 gacaaaccaaagccttaaccacgcctaaagctgaccctgggtggttaataataatggcaaaaggc 4186  
Db 3421 gacaaaccaaagccttaaccacgcctaaagctgaccctgggtggttaataataatggcaaaaggc 3480  
Qy 4187 attgtcattgacagccaaaatggtaaaaataaccatcacagagactaaagcaacctctagct 4246  
Db 3481 attgtcattgacagccaaaatggtaaaaataaccatcacagagactaaagcaacctctagct 3540  
Qy 4247 aatgttaccaatgataaaggtagcgtacgcacacacagacagggcgaataataatacaagac 4306  
Db 3541 aatgttaccaatgataaaggtagcgtacgcacacacagacagggcgaataataatacaagac 3600  
Qy 4307 gaagacaaaacccgtgcgcagcattgttgatgtgtaagcgagggctttaacttgcaa 4366  
Db 3601 gaagacaaaacccgtgcgcagcattgttgatgtgtaagcgagggctttaacttgcaa 3660  
Qy 4367 ggcgaatggtgaagcgttgactttgtctcaacttatgacacgctcaactttgcccagtgcc 4426  
Db 3661 ggcgaatggtgaagcgttgactttgtctcaacttatgacacgctcaactttgcccagtgcc 3720  
Qy 4427 aatgccaccacgctaaggtagacctatgatgacacagcaaaacccagtaaaagtggtctat 4486  
Db 3721 aatgccaccacgctaaggtagacctatgatgacacagcaaaacccagtaaaagtggtctat 3780  
Qy 4487 gatgtcaatgtggatatacaaccattgaagttaaagataaaaacacttggcgtaaaaacc 4546  
Db 3781 gatgtcaatgtggatatacaaccattgaagttaaagataaaaacacttggcgtaaaaacc 3840  
Qy 4547 accacattgaccagttactggcacagggtgctaataaaattgcccctaagcaatcaagctact 4606  
Db 3841 accacattgaccagttactggcacagggtgctaataaaattgcccctaagcaatcaagctact 3900  
Qy 4607 ggcgatgcgcttgcaaggccagtgatcgttgcctcatctaaacacctatctgcccag 4666  
Db 3901 ggcgatgcgcttgcaaggccagtgatcgttgcctcatctaaacacctatctgcccag 3960  
Qy 4667 atccaaactgccaaaggggcgaagccaagcgaacaaactcagcaggtatgttgatgtcgtat 4726  
Db 3961 atccaaactgccaaaggggcgaagccaagcgaacaaactcagcaggtatgttgatgtcgtat 4020  
Qy 4727 ggcgaatagggtcatctatgacagttaccgataaacaagttactatcaagccaaaaatgatggc 4786  
Db 4021 ggcgaatagggtcatctatgacagttaccgataaacaagttactatcaagccaaaaatgatggc 4080  
Qy 4787 acagttgataaaaccaaagaagttgccaaagacaaaactgctgcccaagcccaacccca 4846  
Db 4081 acagttgataaaaccaaagaagttgccaaagacaaaactgctgcccaagcccaacccca 4140  
Qy 4847 gatggcacattggctcaaatgaaatgtaaatcagttcaatacaagaagaacaaagtaaatgat 4906  
Db 4141 gatggcacattggctcaaatgaaatgtaaatcagttcaatacaagaagaacaaagtaaatgat 4200  
Qy 4907 gccaaataaaagcaaggcatcnaatgaagacaaacgcctttgttaaaggacttgaaaaagcc 4966  
Db 4201 gccaaataaaagcaaggcatcnaatgaagacaaacgcctttgttaaaggacttgaaaaagcc 4260  
Qy 4967 gcttctgatacaaaaaacaaaacgcgcagtaactgtgggtgatttaaatgcccgttgcc 5026  
Db 4261 gcttctgatacaaaaaacaaaacgcgcagtaactgtgggtgatttaaatgcccgttgcc 4320

Qy 5027 caaacacgcgtgacctttgcaggggatacagggcacacagcgttaaaaaactggcgagact 5086  
Db 4321 caaacacgcgtgacctttgcaggggatacagggcacacagcgttaaaaaactggcgagact 4380  
Qy 5087 ttgaccattcaaaagtgggcaaacacacacaaataagctaaacgataataaacatcgggttg 5146  
Db 4381 ttgaccattcaaaagtgggcaaacacacacaaataagctaaacgataataaacatcgggttg 4440  
Qy 5147 gtacaggtactgatgcttcaactgtcaaaacttgccaaagacctaaccatcttaacagc 5206  
Db 4441 gtacaggtactgatgcttcaactgtcaaaacttgccaaagacctaaccatcttaacagc 4500  
Qy 5207 gttaatcaggtggcaccacaaattgatgacaaagcgtgtctttttgtagactcaagcgtt 5266  
Db 4501 gttaatcaggtggcaccacaaattgatgacaaagcgtgtctttttgtagactcaagcgtt 4560  
Qy 5267 caagccaaagcaaacacccctgtctaaagtgccaatgggctggacctgggtggcaagctc 5326  
Db 4561 caagccaaagcaaacacccctgtctaaagtgccaatgggctggacctgggtggcaagctc 4620  
Qy 5327 atcagtaagtgggcaaaagcacaaaagataccgacgtgccaatgtatacaacagttaaac 5386  
Db 4621 atcagtaagtgggcaaaagcacaaaagataccgacgtgccaatgtatacaacagttaaac 4680  
Qy 5387 gaagtacgcaacttgttgggtcttggtaagtctgttaataatgataacgctgacggcaaatcag 5446  
Db 4681 gaagtacgcaacttgttgggtcttggtaagtctgttaataatgataacgctgacggcaaatcag 4740  
Qy 5447 gtaaacattgcccagacatacaaaaagaccccaattcaggttcatcatctaaacgcaactgc 5506  
Db 4741 gtaaacattgcccagacatacaaaaagaccccaattcaggttcatcatctaaacgcaactgc 4800  
Qy 5507 atcaaaagcagcagcgttacttggcgttaaaagtaataacgataccgaaaaacttgcact 5566  
Db 4801 atcaaaagcagcagcgttacttggcgttaaaagtaataacgataccgaaaaacttgcact 4860  
Qy 5567 ggtggtatcacagttggcgttgataaaagcgcacgcgttaacggttttaagcaaatgtt 5626  
Db 4861 ggtggtatcacagttggcgttgataaaagcgcacgcgttaacggttttaagcaaatgtt 4920  
Qy 5627 tgggtcaaaaaccccaaaaagatggcagcaaaaagccctgtcgcacttataacgcccga 5686  
Db 4921 tgggtcaaaaaccccaaaaagatggcagcaaaaagccctgtcgcacttataacgcccga 4980  
Qy 5687 ggtcagacccaactatttgaccaaacaccccgacaagcattgacagataataatgaacaa 5746  
Db 4981 ggtcagacccaactatttgaccaaacaccccgacaagcattgacagataataatgaacaa 5040  
Qy 5747 ggtatccgcttcttccatgtcaacgattggcaatcaagagcctgtgttacaagggcgtaac 5806  
Db 5041 ggtatccgcttcttccatgtcaacgattggcaatcaagagcctgtgttacaagggcgtaac 5100  
Qy 5807 ggcattgactcaagtgcctcagggaagcactcagtggcagtaggtttccagggccaagga 5866  
Db 5101 ggcattgactcaagtgcctcagggaagcactcagtggcagtaggtttccagggccaagga 5160  
Qy 5867 gatgtgaagccgcgttggccataggcagacaaacccaagcaggaacccaatccatccgc 5926  
Db 5161 gatgtgaagccgcgttggccataggcagacaaacccaagcaggaacccaatccatccgc 5220  
Qy 5927 atcgggtgataacgcacaaagccacggcgatcaatccatcgcgcattcgttacagggcaatgt 5986  
Db 5221 atcgggtgataacgcacaaagccacggcgatcaatccatcgcgcattcgttacagggcaatgt 5280  
Qy 5987 glacaggttaagcactctggtgccatcggcgacccaagcactgtttaaggctgataacagt 6046  
Db 5281 glacaggttaagcactctggtgccatcggcgacccaagcactgtttaaggctgataacagt 5340  
Qy 6047 tacagttggtgataaacaacacagtttaccgatgccactcaaacagatgtctttggtgtg 6106  
Db 5341 tacagttggtgataaacaacacagtttaccgatgccactcaaacagatgtctttggtgtg 5400





Db	44767	gacacaaatacatcaaaaggtgtcctcaacactgactgacattaccacacagat	44826
Qy	2951	agtagcgcaacatagaactggygcatacaatccaaagacaaagacaaatccacagctgcc	3010
Db	44827	aatacacacacacagacaagaacaatgacgcccaagggcaag--aaaccacagctgcc	44883
Qy	3011	agcattaatgatataataacacagcgttttaacctaaactaaaaataataaacacccccattgac	3070
Db	44884	agtaacaaagatgctaaacagcaggttttaactgcaaggcaatgycgaagccgttgac	44943
Qy	3071	tttgtctccacttatgacattgttgactgttgcaattggcaattgcaacacccacagta	3130
Db	44944	tttgtctacttatgataccgtcaactttgccaatggcgatggccacctgcccagtt	45003
Qy	3131	acc-----catgataccgctcaacaaacacagtaaaagtgtgtatgatgtgaattggtgat	3184
Db	45004	acctataattataatgaaggcaaacacacagcacagctgtctataatgtcaaatgtggat	45063
Qy	3185	gatacaacattcatcttaacaggccactgtagacataataaaaaacttggctcaaaaacacc	3244
Db	45064	gatacaacattcatctgccaataataatgtgtaataaaaaacttagcgtaaaacacacc	45123
Qy	3245	aaactgaacaaaaaagtgctaattgtaatacacagcaactaaacttaattgttaacttagt	3304
Db	45124	acattgcaaaaaaacaatggtgcaaatggttaatgcaactaaatttagtgcaac-----	45177
Qy	3305	gatgaagatgacctgtttaacgcgaagacatcgccgaaatacttaaacacaccttagccaag	3364
Db	45178	aatggcgatgacctgtttaacgcgaagatgttccgcacaatcacaacatccttagccgag	45237
Qy	3365	gaattcacacacaaagacagacagcagcagccctcacaaaccttaccogttaaaaag	3424
Db	45238	gaattcacacacaaagacagcagcagcagcagccctcacaaacctttaaagtcaaaaa	45297
Qy	3425	gtagatgaataataatgtgatgacgccaacgcacacggtgggtgcaaaagaaagcga	3484
Db	45298	gacggtgcaaccaatgatgatg-----acaccaacacgctgggt-----	45336
Qy	3485	aataatacaagatcaacaccttaacactcaaaagtgaaaacggtcttaattataaacacgac	3544
Db	45337	aaaaatgctgacacaccttaacattcaaaagtgaaaacggtctcaacggttgcctaccaaa	45396
Qy	3545	aaaaatgctgacggttaccttggcattaaacacacacacggtctttaaagcggg-----caaa	3601
Db	45397	aaagatggtacggttaacctttggcattacacccccaaagcggctttaaagcggcgacaac	45456
Qy	3602	agacccctaaacgacggtggtctgtctattaaaaacccccactgggtagcgacaataccaa	3661
Db	45457	accaactaaacaaagatggtctgtctattaaaaacccccactgggttaacgaaacaaatccaa	45516
Qy	3662	gacggtgctgatggcgtaagtttgcgaagtttaataataatgggtgtgttaggtgctggc	3721
Db	45517	gtcgggtgctgatggcgtaagtttgcgaagtttaataataatgggtgtgttaggtgctggc	45576
Qy	3722	attgatggcaaacctgcattacacagataagaaattgggtttactgggactaatggctca	3781
Db	45577	attgatggcaaacctgcattacacacagataagaaattgggtttactgggactaatggctca	45636
Qy	3782	cttgataaaagcaaacccccacttaagcaagacggttaacgaaggtggttaaaaagatt	3841
Db	45637	cttgatacaacaaacccccacttaagcaagacggttaacgaaggtggttaaaaagatt	45696
Qy	3842	accaacattcaatcaggttgagattgcccacaaacacgcatgatctgtgacagcgccgaag	3901
Db	45697	accaacattcaatcaggttgagattgcccacaaacacgcatgatctgtgacagcgctcag	45756
Qy	3902	atttatgtttaaaaacacgaacttgaacaaacaaatcatcagcagctgcccacacagcacaa	3961
Db	45757	atttatgtttgaacacgaacttgaacaaacaaatcatcagcagctgcccacacagcacaa	45816
Qy	3962	aactcattacagaattctcagtagcagatgaacaaaggttaataactttaacggttagtaac	4021
Db	45817	aactcattacagaattctcagtagcagatgaacaaaggttaataactttaacggttagtaac	45876
Qy	4022	ccttactccagtttatgacacctcaaaagaccttgatgctcatcacctttgcaggtgaaaaac	4081
Db	45877	ccttactccagtttatgacacctcaaaagaccttgatgctcatcacctttgcaggtgaaaaac	45936
Qy	4082	ggcattaccacccaaggttaataaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	4141
Db	45937	ggcattaccacccaaggttaataaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	45996
Qy	4142	ttaaccacgctgaactgacggtgggttaataataatggtgcaaaaggtctgtcatgtgacagc	4201
Db	45997	ttaaccacgctgaactgacggtgggttaataataatggtgcaaaaggtctgtcatgtgacagc	46056
Qy	4202	caaaatggtcaaaataccatcacaggactaaagcaaacacttagctaatgtttaccaatgat	4261
Db	46057	aaagtgtgcaaaataccatcacaggactaaagcaaacacttagctaatgtttaccaatga-	46115
Qy	4262	aaagtgtgctacgacacacagaagggcaataataataaagcgttgcgataaacacggacaaaac	4321
Db	46116	-----tggtgcagacacgacacgaagccaaagggcttgcgataaacacggacaaaac	46170
Qy	4322	gcccgcagcattgtgtgatgtgctgaagcgcagggctttaaacttgaaggcaatgtgtgaagcg	4381
Db	46171	gctgcacgactgt	46230
Qy	4382	gttgacttgtctcacttatgacacgctcaacttgcggtggaatgcaacacacgct	4441
Db	46231	gttgacttgtctcacttatgacacgctcaacttgcggtggaatgcaacacacgct	46290
Qy	4442	aaggtgacctatgtgacacagaacaaacccgtaaggtgtgtctatgtatgtatgtgtgtgtgt	4501
Db	46291	aaggtgacctatgtgacacagaacaaacccgtaaggtgtgtctatgtatgtatgtgtgtgtgt	46350
Qy	4502	gatacaacatttgaagttaaagataaaaaacttggcgttaaaaaacacacacacacacacacac	4561
Db	46351	gatacaacatttgaagttaaagataaaaaacttggcgttaaaaaacacacacacacacacacac	46410
Qy	4562	actggcacaggtgtgctaataatttcccttaagcaatcaagctacttgcgacgtgcgctgtgtc	4621
Db	46411	actggcacaggtgtgctaataatttcccttaagcaatcaagctacttgcgacgtgcgctgtgtc	46470
Qy	4622	aagccagtgatcgtgtgtctcatctaaacaccttatctggcgacatccaaactgccaacaa	4681
Db	46471	aagccagtgatcgtgtgtctcatctaaacaccttatctggcgacatccaaactgccaacaa	46530
Qy	4682	ggggcagcccaagcgaacactcagcaggtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	4741
Db	46531	ggggcagcccaagcgaacactcagcaggtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	46590
Qy	4742	tatgacagtcacagataacacagttactatcaagcccaaaatgatgtgcacagctgtgataaac	4801
Db	46591	tatgacagtcacagataacacagttactatcaagcccaaaatgatgtgcacagctgtgataaac	46650
Qy	4802	aaagaagtgtcccaaaagacaacttgggtgcccaagcccaacccacagctgtgacattggct	4861
Db	46651	aaagaagtgtcccaaaagacaacttgggtgcccaagcccaacccacagctgtgacattggct	4671

QY 5102 gggcaacagacacccaataaagctaacgataataaacatcgggtggtgtagcaggtactgat 5161  
|||||  
Db 46951 gggcaacagacacccaataaagctaacgataataaacatcgggtggtgtagcaggtactgat 47010  
QY 5162 ggcctcactgtcaaacctgcgaagacgaactaacgaacttaacagcgttaatacgaagtgcc 5221  
|||||  
Db 47011 ggcctcactgtcaaacctgcgaagacgaactaacgaacttaacagcgttaatacgaagtgcc 47070  
QY 5222 accaaattgatgacaaaggcgtgtctttgtgactcaagcgtcaagccaaagcaaac 5281  
|||||  
Db 47071 accaaattgatgacaaaggcgtgtctttgtgactcaagcgtcaagccaaagcaaac 47130  
QY 5282 accctgtgtaagtgcgaatggcctggagcctgggtggaaggttcacgaatgctgggc 5341  
|||||  
Db 47131 accctgtgtaagtgcgaatggcctggagcctgggtggaaggttcacgaatgctgggt 47190  
QY 5342 aaaggcacaagaatccgacgtgcgaatgtacaacagttacaacgaagtacaacgaacttg 5401  
|||||  
Db 47191 aaaggcacaagaatccgacgtgcgaatgtacaacagttacaacgaagtacaacgaacttg 47250  
QY 5402 ttgggtcttgtaagtgttaataacgtgacgcgaactcaggtaaacatctgcgcac 5461  
|||||  
Db 47251 ttgggtcttgtaagtgttaataacgtgacgcgaactcaggtaaacatctgcgcac 47310  
QY 5462 atcaaaaagacccaataatcaggttcacatctcaacgcgactgtcaataagcaggcagc 5521  
|||||  
Db 47311 atcaaaaagacccaataatcaggttcacatctcaacgcgactgtcaataagcaggcagc 47370  
QY 5522 gtacttgccgttaaggtataacgataaccgaataaaccttccactgtgtgatacaagt 5581  
|||||  
Db 47371 gtacttgccgttaaggtataacgataaccgaataaaccttccactgtgtgatacaagt 47430  
QY 5582 ggcgtgataaagcgcgaacgctaacgcggttttaagcaatgtttgggtcaaacccaa 5641  
|||||  
Db 47431 ggcgtgataaagcgcgaacgctaacgcggttttaagcaatgtttgggtcaaacccaa 47490  
QY 5642 aaagatggcagcaaaaagccctcgtcgcaactataacgcgcgaggttcagacaaactat 5701  
|||||  
Db 47491 aaagatggcagcaaaaagccctcgtcgcaactataacgcgcgaggttcagacaaactat 47550  
QY 5702 ttgccaacaccccgagaaagcctatgacagaataatgaacaggtatccgcttttc 5761  
|||||  
Db 47551 ttgccaacaccccgagaaagcctatgacagaataatgaacaggtatccgcttttc 47610  
QY 5762 catgtcaacgatggcaatcaagagcctgtgtacaagggcgtaacgcgcatctgaactcaagt 5821  
|||||  
Db 47611 catgtcaacgatggcaatcaagagcctgtgtacaagggcgtaacgcgcatctgaactcaagt 47670  
QY 5822 gcctcaaggcaagcactcagtggcgataaggtttccagcccaaggcagatggtgaagccgcc 5881  
|||||  
Db 47671 gctgctggcaagcattcagtggcgataaggtttccagcccaaggcagatggtgaagccgcc 47730  
QY 5882 gttgccatagcagacaaaacccagcaggaacccaatccatccatcgcatcggtgataacgcga 5941  
|||||  
Db 47731 gttgccatagcagacaaaacccagcaggaacccaatccatccatcgcatcggtgataacgcga 47790  
QY 5942 caagccacggcgcatcaatccatcgcatcggtgatacaggaacgaatgtggtagcaggttaagcac 6001  
|||||  
Db 47791 caagccacggcgcatcaatccatcgcatcggtgatacaggaacgaatgtggtagcaggttaagcac 47850  
QY 6002 tctggtgccatcgcgaccccaagcactgtttaaggtcgtataacagttacagtgctgggtaat 6061  
|||||  
Db 47851 tctggtgccatcgcgaccccaagcactgtttaaggtcgtataacagttacagtgctgggtaat 47910  
QY 6062 aacacaggtttaccgatgccactcaacccgatgtctttggtgggcaataacatcac 6121  
|||||  
Db 47911 aacacaggtttaccgatgccactcaacccgatgtctttggtgggcaataacatcac 47970  
QY 6122 gtgaccgaaaagtaactcgggttccttaggttcaaacctcgcacatcagtgagggcacacac 6181  
|||||  
Db 47971 gtgaccgaaaagtaactcgggttccttaggttcaaacctcgcacatcagtgagggcacacac 48030

QY 6182 gcaggcacacaagcccaaaaaatctgaaggcacagcaggtacaacacacacacagcaggtgca 6241  
|||||  
Db 48031 gcaggcacacaagcccaaaaaatctgaaggcacagcaggtacaacacacacacagcaggtgcc 48090  
QY 6242 accggtacggttaaaggcttctgctggacaaacgcggttctgctgctcctcgtgggtgcc 6301  
|||||  
Db 48091 acaggtacggttaaaggcttctgctggacaaacgcggttctgctgctcctcgtgggtgcc 48150  
QY 6302 tcaggtgctgaacgccgttatccaaaatgtggcagcaggtgaggtcagtgccaccagcagcc 6361  
|||||  
Db 48151 tcaggtgctgaacgccgttatccaaaatgtggcagcaggtgaggtcagtgccaccagcagcc 48210  
QY 6362 gatcggttcaatgtagccagttgtacaagaaccccaaaacattgccaacgcaaccaat 6421  
|||||  
Db 48211 gatcggttcaatgtagccagttgtacaagaaccccaaaacattgccaacgcaaccaat 48270  
QY 6422 gagcttgaccatcgtatcccaaaacgaataaagcccaatgaggtgaggtatctcagcag 6481  
|||||  
Db 48271 gagcttgaccatcgtatcccaaaacgaataaagcccaatgaggtgaggtatctcagcag 48330  
QY 6482 atggcgatggcgctccatgcccacaagcctacattcctggcagatccatggttaccgggggt 6541  
|||||  
Db 48331 atggcgatggcgctccatgcccacaagcctacattcctggcagatccatggttaccgggggt 48390  
QY 6542 attgccaccccaacgctcaagggtgcggtgcagtgaggactgcgaagctgcggataat 6601  
|||||  
Db 48391 attgccaccccaacgctcaagggtgcggtgcagtgaggactgcgaagctgcggataat 48450  
QY 6602 ggtcaatgggttatataaatcaatggttcagccgatacccaagccatgtagggcgca 6661  
|||||  
Db 48451 ggtcaatgggttatataaatcaatggttcagccgatacccaagccatgtagggcgca 48510  
QY 6662 gttggtgcaggttttctacttttaagccataaatgcgaagttttacttaaaaaatcaatct 6721  
|||||  
Db 48511 gttggtgcaggttttctacttttaagccataaatgcgaagttttacttaaaaaatcaatct 48570  
QY 6722 caccatggtgtataaaaaacagcatacagcatcagtcataatctactgatgctgatgttttta 6781  
|||||  
Db 48571 caccatggtgtataaaaaacagcatacagcatcagtcataatctactgatgctgatgttttta 48630  
QY 6782 tcacttaaacattttaccgctcaagtgatctctttcaccatgccaataatgcgcatgga 6841  
|||||  
Db 48631 tcacttaaacattttaccgctcaagtgatgtgtcttccaccatgccaataatgcgcatgga 48690  
QY 6842 tcataggttaaaacttattgagtaaaattttatcaatgtagttgtagatatggttaaaattg 6901  
|||||  
Db 48691 tcataggttaaaacttattgagtaaaattttatcaatgtagttgtagatatggttaaaattg 48750  
QY 6902 tgccattgacaaaataatgacccgtttatcccgaaaattttotgattatgacgttgacc 6961  
|||||  
Db 48751 tgccattgacaaaataatgacccgtttatcccgaaaattttotgattatgacgtttttta 48810  
QY 6962 tgcaggtc 6969  
|||  
Db 48811 taaatttc 48818  
|||  
RESULT 8  
AAF59104  
ID AAF59104 standard; DNA; 6159 BP.  
XX  
AC AAF59104;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE M. catarrhalis strain Q8 200kDa gene SEQ ID NO:8.  
XX  
KW Moraxella catarrhalis strain Q8; major outer membrane protein;  
KW 20kDa outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection; ds.  
XX  
OS Moraxella catarrhalis.  
XX







Qy	2069	ttggataaaaaaacacttaaaagtgggttagttgcaattaccatagacaatggcattgat	2128
Db	1513	cttgacaacgaagaagctaaagtgtggcaacgagccctaaacagtggttagctgactgtt	1572
Qy	2129	gcaggtaataaaaaagatcagtaactctgccaaaggtagcagtgctaacgagtcgggttacc	2188
Db	1573	aataacacactggttaataaaaaacatccaagtcggtgctaatggcattaaatttgcaca	1632
Qy	2189	atcgaaacagctcaaaagccgccgaagctacttttaacgcaggcgtgcgtccatcagtgca	2248
Db	1633	gtcgtaataatgttgcataatcctcagcaacagtcggcactgctgtattaccgaagag	1692
Qy	2249	ccctactgaaatatcagttgatgctaaagatggcaatgttaccgcccaacttacaaca--	2306
Db	1693	aaatttggtttgctggtactaatgatggatgtgatgaacagcaccatatttggataaa	1752
Qy	2307	-----ttggcgtgaaacaccacgagcttaacagtgatggcactagtgtataaaattagt	2359
Db	1753	gaacgacttaaaagtgggtcggtgtgaattaccacagatagggtattaaagctggtaat	1812
Qy	2360	gttaaggtagtggtagcaacaatagcttagttaccgccgaacatttggcaagctatcta	2419
Db	1813	cacaagattaccggacttactaatgtatagcaaataccgatccggttaccatcaaacag	1872
Qy	2420	aatgaagtaaatcgaaacgctgaacagtgctctacaaagctttaccgtttaagaagaagac	2479
Db	1873	ctcaagagcccaagcctactttaacgcagcggatggcattcagttatataagtaataac	1932
Qy	2480	gatgatgaagc-----ccaaagctatccacgctggct	2509
Db	1933	gggagtctagtgtagtagtggaatataccaccaactttaacattagcgtgaaa	1992
Qy	2510	aaagatacagacaaaaaattccggccagtcagcatctttaaactcaaggttaaaaacggt	2569
Db	1993	accactaagcttaacagtaatggcaccagtggaataataaaatttagtggtagtaatgct	2052
Qy	2570	ctaacggtgctacaaaaaagatgtagcgggttacctttgggttagccgaagatagcgtt	2629
Db	2053	catgataacaatagcttagttaccgccaaagatttggcagactctcaataaagtcgat	2112
Qy	2630	ctgacctatggcaaaagcaccctaaacaacgatggcttgactgttaaagataccaacgaa	2689
Db	2113	gaaacggctgacagtctctaccaagcittaaagtccaaacggtgataatgcaacaac	2172
Qy	2690	caaatccaagtcggtgcttaatggcattaaattactaatgtgaatggttagtaatccaagtt	2749
Db	2173	gccatcacggtgggttaagatacaaacgcgaagaccttcaacaccttaaaactcaagggt	2232
Qy	2750	actggcatgcaaataccgctcgatla---ccagagataaaattgcttttgcgttct	2806
Db	2233	gaaacacgggtgtaattatcgaccaatagcagcaggtacagttacctttggcattgac	2292
Qy	2807	gatggtgcagttgatac-----aaacaaaccttacttgcataagacaagctla--caa	2857
Db	2293	caaaagtaatggtctcccacgcttaagctgacgggtgggtagcgatacaaatggtaatcga	2352
Qy	2958	gttggcaatgttaagattaccaaacctggcataacgcaggtgtgtaagccatccacaggg	2917
Db	2353	ttggttattgacaagtcctcagctgcggttaacagcacaacaaacatcatttaaagga	2412
Qy	2918	ctgtcccaacactgctcagcattccgatcaaatagccgaacatagacatggccaat	2977
Db	2413	ttgtccccaacactgctcagctatccagtcgaagtgccgaacatagcactgggcaat	2472
Qy	2978	acaatccaagacaagaacaaatccaacgctgcagcattatgatataattaaatacagcc	3037
Db	2473	acaatcgaaacaaacaaatccaacgctgccagattgatgatgctgctaaatlgcagcc	2532
Qy	3038	tttaacctaaaaataatacaaacccattgactttgtctccacttatgcaattgttgac	3097
Db	2533	tttaacctaaaaataatggcaagacaaagacttggctccactatlgacactgttgac	2592

QY	3098	tttgccaatggcaatgccaccaccgcgcaacagtaaacccatgatccgctcaacaaacacagt	3155
DB	2593	tttatcgatggcaatggccaccaccgcgcaacagtaaccttatgatgaagccaatcaaacctgt	2652
QY	3158	aaagtgggtatatgatgtgaatgggtgatatacaaaccatcatcatcaacggcactgatgac	3217
DB	2653	aaagtggcgctatgatgtgaatgggtgatatacaaaccatcatcatcaacggcactgatgac	2712
QY	3218	aataaaaaacttggcgctaaaaaccaccacaaactgaacaaaacgaagtgtcaatggttaataca	3277
DB	2713	aagaaacaacttggcgctaaaaaccatcaaaactgacgcgaacaagtactaatggttaatgca	2772
QY	3278	gcaactaaactttaatgttaactctagtatgaagaatgcccctgtttaacggccaaagacatc	3337
DB	2773	ac-----tacatttagtacgcagatgacccatgcccctgtttaagccagtgatcat	2823
QY	3338	gccgaaaatctaaacaccctagccaaggaatactcacaccacaaaggcacagacaccc	3397
DB	2824	gcggcaatctaaacaccctagccggggaatactcacaccacaaaggcacagacacacc	2883
QY	3398	gccctacaaacctttaccgttaaaaaggtagatgaaaaataataatgctgatgcgcgaac	3457
DB	2884	gccctacaaacctttaccgttaaaaaggtagatgaaaaataataatgctgatgcacccaac	2943
QY	3458	gccatcccggtgggtccaaagaacgcgaataataatcaagtcaacacctcaacccaaggt	3517
DB	2944	gccatcccggtggtaaaagatggcacaagtggttaaaagtcacacaccttaaaactcaaaagt	3003
QY	3518	gaaacccggtcttaataattaaaaccgcgaataaatggttacgtttacccttggcattaacacc	3577
DB	3004	aaaaacccggtcttgatattaaaaccgcgaagaatggttacggttacccttggcattaacacc	3063
QY	3578	acaagccggtcttaaaagccggcaaaagcacac---ctaaacgacggtggcttgcctattaaa	3634
DB	3064	caaaagccggtcttaaaagccggcagacacacacctcttaaaacaaatggcttgcctattaaa	3123
QY	3635	aaccocactggtagcgaaacaaatcccaagtcggtgctgtgatggcgtaagtttgcgaaggtt	3694
DB	3124	aaacccggttagtaaacgaacaaatcccaagtcggtgctgtgatggcgtaagtttgcctatggt-	3182
QY	3695	aataataatgggttttagtgtctggcattgtatggcaacactgcgattaccagagatgaa	3754
DB	3183	--taataatgggttttagtgtctggcattgtatggcacaactgcgattaccagagatgaa	3240
QY	3755	attggcttactgggactaatggctcacttgataaaagcaaacccccactcaagcaaaagac	3814
DB	3241	attggcttactgggactaatggctcacttgataaaagcaaacccccactcaagcaaaagac	3300
QY	3815	ggcattaacgcaggttggttaaaaagattacaacatcatcaggtgagattgcccacaaac	3874
DB	3301	ggcattaacgcaggttggttaaaaagattacaacatcatcaggtgagattgcccacaaac	3360
QY	3875	agccatgatgctgtgacagggcggaagatttatgatttaaaaaacccaacttgaaaacaaa	3934
DB	3361	agccatgatgctgtgacagggcggaagatttatgatttaaaaaacccaacttgaaaataaa	3420
QY	3935	atcagcagttactgccaacaaacagcacaaaaactcatcacgaaattctcagtagcagatgaa	3994
DB	3421	atcagcagttactgccaacaaacagcacaaaaactcatcacgaaattctcagtagcagatgaa	3480
QY	3995	caaggttaataactttacggttagtaaccccttactcagtttatgacacctcaaaagacctct	4054
DB	3481	caaggttaataactttacggttagtaaccccttactcagtttatgacacctcaaaagacctct	3540
QY	4055	gatgtaatacactttgcaggtgaaaaacgcgcaattaccaccaaggttaataaaggttggtgt	4114
DB	3541	gatgtaatacactttgcaggtgaaaaacgcgcaattaccaccaaggttaataaaggttggtgt	3600
QY	4115	catgtgggcattgacaaacccaaggttaaacacgccttaagctgacctgggtaataat	4174
DB	3601	catgtgggcattgacaaacccaaggttaaacacgccttaagctgacctgggtaataat	3660
QY	4175	aatggcaaaaggaattgtcattgtgacagccaaaatgtgtcaaaataaccatccacagactaaac	4234





1882	gacggtgccctacaa-----agctctctctattctgtagtaaaaaaggtcaaggaa	1929
Db		
2636	attggcaaaagcacctaaacacagatggcttgactgttaaagatacccaacgaacaaatc	2695
Qy		
1930	tttacgattagtaacttgattcttaattggttaataccccaataacacctttgagaccatcc	1989
Db		
2696	caagtcggtgtaattggcattaaatttactaattgaaatggtagtaattccaggtactggc	2755
Qy		
1990	tttgcaaggtagaaaagcgcatcgatcagcaatgacatgacgaaggttaaaatcgaag--	2047
Db		
2756	attgcaaaataccgctcgcatcaccagagataaaaattggcttttgctggtttctgatggtgca	2815
Qy		
2048	-ttgctattgccccaatcaatggctcaccacgcttaagctgacctgggtagcgataaa	2106
Db		
2816	gttgatacaaaacaaaccttatcttgatcaagacaagctacaaagtggcgaatgttaagatt	2875
Qy		
2107	gatggtataaactcaattggtttattggagaag-----	2137
Db		
2876	accaacactggcaattaacgycaggtggtataagccatcacagggctgtccccaacacgtgcct	2935
Qy		
2138	-----tggtctaggggttaacgacacccaacaaacatcatlagaggatgtgccccaacacgtgcct	2193
Db		
2936	agcatgccgatacaaaagtagc---cgcaacatagaaactgggcgaataacaaatccaa	2989
Qy		
2194	agcattaccaatgcggtggctacgcaccacagacagggcgaatacacaatcaaccagcgac	2253
Db		
2990	aaagcaaaatccaaacgctgccagcattaatgatatattataacacaggtcttaacctataaa	3049
Qy		
2254	gaagcaaaatcccaaaagccgcagtatcggtgatatataataacagggctttaacctataaa	2313
Db		
3050	ataataaacaccccatgactttgtctccacttatgacattgttgactttggcgaatgggc	3109
Qy		
2314	ataataagaacitccgttggtttgtctccacttataacactgttgactttatcgatggc	2373
Db		
3110	aatggccaccgcgcacagtaacccatgatacccgataacgctaaacacacagtaagtggtatat	3169
Qy		
2374	aatggccaccgcgtaaggtaacttaacgtgaacacaaatcaaacacagtaagtaacttat	2433
Db		
3170	gatgtgaattggtatgatacaaacatcatctaacagggcactgtagcaaa---taaaaaa	3226
Qy		
2434	gatgtcaattggtatgagaaaacccatggaactcacagggcgataatggcagacacaaacaa	2493
Db		
3227	cttgccgtcaaaaaacacaaactgaacaaacaaagtgctaattggttaatacagcaactaac	3286
Qy		
2494	attggcgtcaaaaaacacacacactgaccacaacaaatgctaattggtataaagcaac-----	2546
Db		
3287	tttaatgttaactctagtgtgaagatgcccttgttaacgcgaagacatcgccggaataat	3346
Qy		
2547	--caactttagaccacgataacgatgccttgttaacgcgaagacatcgccggaataat	2604
Db		
3347	ctaaacaccttagccaaggaaattcaacaccacaaaggcacagcagacaccccgctacaa	3406
Qy		
2605	ctaaacaccttagccaaggaaattcaacaccacaaaggcacagcagacaccccgctacaa	2664
Db		
3407	acctttacgttataaaaggttagtgaaaataataatgctgtgacgcgcgaacgcgaatcacc	3466
Qy		
2665	acctttaagatcaaaaaaagacggtgca----actgatgacgaacacatcaccggtgggt	2718
Db		
3467	gtgggtcaaaagaaacgcaaaataataacgttaacaccccttaacacactcaaaaggtgaacacgt	3526
Qy		
2719	aaagatggtcacaaaaacggaagaccgtcaacacactcaaaactcaaaaggtgaacacgt	2778
Db		
3527	cttaataataaaacgcgaataaattggtacggtttacgtttggcatataacacacacagcggt	3586
Qy		
2779	ctaacggttgctaccaaataaagatggtacggtttacctttggcatataacacacacagcggt	2838
Db		
3587	cttaaaagccggcaaaagcacc---ctaaacgcggtggtctgtctattataaaaccccat	3643
Qy		
2839	cttaaaagccggcgacagcaccactctaaacaaagatggcttgctattataaaaccccggt	2898
Db		
3644	ggtacgcaacaaatccaagtcggtctgatggctgaagtttggcaaggttaataataat	3703
Qy		
2899	agtaacgaacaaatccacgctgggtgctgatggcggtgaagtgttggcaaggt---tgataag	2955
Db		

Qy	3704	gggtgtgtagggtgctggccattgatggccacaactgcattaccagaagatgaattggcttt	3763
Db	2956	ggtaattcaagcactggcattgatggcaagcggtatccaccaagatcaaatggcttt	3015
Qy	3764	actgggacctaaatggctcacttgtataaagacaaacccacccttaagcaaacg	3814
Db	3016	actggggctaaatggctcacttgtataccaccaaaacccacccttaaccaaagacagcttaaa	3075
Qy	3815	-----ggcattaacgcaggtgtaaaaaagattaccaac	3847
Db	3076	gtgggtgaagttgaaattaccaacactggcattaacgcaggtgtgtataaagattaccaac	3135
Qy	3848	attcaatcagggtgagattgcccataaacagccatgatgctgtgacagcgcgcaagatttat	3907
Db	3136	attcaatcagggtgattatcccaaacagcaatgatgctgtgacagcgctggtgttat	3195
Qy	3908	gatttaaaacccgaacttgaatacaaaaatcagcagctactgccaaaaacagcacaaaaactca	3967
Db	3196	gatttaaaacccgaacttgaagacaaaatcaacagtgctgtctaaaaacagcacaaaaactca	3255
Qy	3968	ttacacgaattctcagtagcagatgaacaaggtataactttacggttagtaacccttac	4027
Db	3256	ttacacgaattctcagtagcagatgaacaaggtaatcactttacggttagtaacccttac	3315
Qy	4028	tcagttatgacaacctcaaaagacctctgatgtcatcactcttgccaggtgaaaaacggcatt	4087
Db	3316	tcagttatgacaacctcaaaagacctctgatgtcatcactcttgccaggtgaaaaacggcatt	3375
Qy	4088	accacaaggtataataaagtggtgctgctggggcattgaccaaaccaaaagccttaacc	4147
Db	3376	accacaaggtataataaagtggtgctgctggggcattgaccaaaccaaaagccttaacc	3435
Qy	4148	acgcctaagctgacogtgggtataataatggcaagggcattgtctattgacagccaaaat	4207
Db	3436	acgcctaagctgacogtgggtataataatggcaagggcattgtctattgacagtaaaagt	3495
Qy	4208	ggfcaaaatcacctcacaggaactaagcaaacactctagctaagttaaccaatgataaaggt	4267
Db	3496	ggfcaaaatcacctcacaggaactaagcaaacactctagctaagtgtaccaatga-----t	3549
Qy	4268	agcgtacgcaccacagaaacagggcaataataatcaaaagcagaagacaaaacccgtgcgc	4327
Db	3550	ggfgcagacagcactaagcgaaggcctggccaatgacacccgacaaaacccgtgcgc	3609
Qy	4328	agcattgtgtatgtctaaagcagggctttaaactgcaaggaatgtgtgaagcggttgac	4387
Db	3610	agcattgtgtatgtctaaagcagggctttaaactgcaaggaatgtgtgaagcggttgac	3669
Qy	4388	tttgtctccacttatgacaccgtcaactttgccatggcaatgccaccacgcgtcaaggtg	4447
Db	3670	tttgtctccacttatgacactgtgactttaatgataggcaatgccaccacgcgtcaaggtg	3729
Qy	4448	acctatgagacacaagcaaaacccagtaaaagtggtctatgatgtccaatgtggatgataca	4507
Db	3730	acctatgagacacaagcaaaacccagtaaaagtggtctatgatgtccaatgtggatataaa	3789
Qy	4508	accattgaagttaaa---gataaaaaacttggcttaaaaaacccacacatttgaccagttact	4564
Db	3790	accattgaagttgacaagttgataaaaaacttggcttcaaaaacccacacacattgacaaaaca	3849
Qy	4565	ggcacagggtcctaataaaattggcctaagcaatcaagctactgcgcgtgcgtctgtcaag	4624
Db	3850	agtgtctaattggttaatgca---accataatttagtgcgcgcggtggcgatggccttgttaaa	3906
Qy	4625	ggcagtgatatcgttgtctatcaataaaccttattctgggcacttccaaaactgccaaaggg	4684
Db	3907	ggcagtgatatcgcaccacatctaaataaccttggctggcgacatccaaaacgcgcaaaagg	3966
Qy	4685	gcaagccaagcgaaacaactcagcaggttatgtgatgctgtatgcaataagggtcatctat	4744
Db	3967	gcaagccaagcagcgtcagcaggttatgtgatgctgtatgctgtatgcaataagggtcatctat	4026

QY	4745	gacagtaccgataaacaagctactatcaagccaaaatgtatggcacagttgataaaacaaa	4804
DB	4027	gcagttaccgataaagaagttactatcaagtcaatgacaaaggtcaagtggacaaaacaaa	4086
QY	4805	gaagttgccaaagacaaaactggtcgcccaagcccaaaccccaatggcacattggctcaa	4864
DB	4087	gaagttgccaaagacaaaactggtcgcccaagcccaaaccccaatggcacattggctcaa	4146
QY	4865	atgaattgctcaaatcagttcatcaacaagaacagtaataatgattgcctaataaaagcaagc	4924
DB	4147	atgaattgctcaaatcagttcatcaacaagaacagtaataatgattgcctaataaaagcaagc	4206
QY	4925	atcaatgaagacaaacgcctttgttaaaggacttgaaaaagccgcttctgtgatacaaaaacc	4984
DB	4207	atcaatgaagacaaacgcctttatcaaaaggttgaaaaagccgccaagaacacccaaaacc	4266
QY	4985	aaaaagccgcagtaactgtgggtgatttaaatgcgcgttgcccaaacacgcgtgacctttt	5044
DB	4267	aaaaagccgcagtaactgtgggtgatttaaatgcgcgttgcccaaacacgcgtgacctttt	4326
QY	5045	gcaggggatacagggcacacgcgttaaaaaactggcgagactttgaccttcaaaagtgggg	5104
DB	4327	gcaggggatacagggcacacgcgttaaaaaactggcgagactttgaccttcaaaagtgggg	4386
QY	5105	caaacagacacacaataaagctaacccgataataacatcgggtgtgtagcagtgactgatgc	5164
DB	4387	caaacagacacacaataaagctaacccgataataacatcgggtgtgtagcagtgactgatgc	4446
QY	5165	ttcactgtcaaaacttgccaaagaccttaacaaatctttaacagcgttaatgcagttggcaac	5224
DB	4447	ttcactgtcaaaacttgccaaagaccttaacaaatctttaacagcgttaatgcagttggcaac	4506
QY	5225	aaaaattgatcaaaaagcgtgtctttttgtagactcaagcgggtcaagcccaagcaaacacc	5284
DB	4507	agaattgatgaaaaagcgtgtctttttgtagacgcaaacggtcgaagcccaagcaaacacc	4566
QY	5285	cctgtgctaaagtgcgaatgggctggacctgggtgggc-----	5320
DB	4567	cctgtgctaaagtgcgaatgggctggacctgggtgggc-----	4626
QY	5321	-----	5320
DB	4627	gctgttgatgataacgatcggtgaactttaacagactttaatgaagttgccaaaacggtc	4686
QY	5321	-----	5320
DB	4687	aaacacctaaacaaacaaagtaactcaggtgcgtctattaccctttgtgtaaacgtagcc	4746
QY	5321	-----	5320
DB	4747	aatggcaagcccatcaatggccacgatggcaagcccaaaaagccatcaagggcgccgat	4806
QY	5321	-----	5320
DB	4807	ggtaaatactatcacgcccaacgcgaacgcgtacctgtggacaaagatggcaagcccatc	4866
QY	5321	-----	5320
DB	4867	acagatgcggacaaaacttgccaatctggcagctcattggcaaaccccttgatgcaggtcat	4926
QY	5321	-----	5320
DB	4927	caagttgtggcaagcctagcggcgaactcagatggccatcacccctaaccaacatcaagtcc	4986
QY	5321	-----	5320
DB	4987	actttgccacaaattgacacacacaaacacaggttaatgcgaatgcagggccaaagcccaagt	5046
QY	5321	-----	5320
DB	5047	ctgccagcctatcagcagcagcaagaagttaatgctgcagtgctcaagaatgctgctaaat	5106
QY	5321	-----	5320

Db	6187	atcgggtgataacgcaacagccacggcgagatcaatccatcgcactcgggtacaggccaatgtcg	6246
Qy	5987	gtacaggtaagcactctgggtgccatcggcgacccaagcactgttaaggtcgtataacagt	6046
Db	6247	gtaacaggtaagcactctgggtgccatcggcgacccaagcactgttaaggtcgtataacagt	6306
Qy	6047	tacagtgtgggttaatacaaacaccagttaccgaatgccactcaaacagatgtctcttggtcg	6106
Db	6307	tacagtgtgggttaatacaaacaccagttatcgatgccactcagaccgagtcctctgggtcg	6366
Qy	6107	ggcaataacaatacacctgacccgaagaataactcgggtgccttaggttcaaaactctgcatac	6166
Db	6367	ggcaataacaatacacctgacccgaagaataactcgggtgccttaggttcaaaactctgcatac	6426
Qy	6167	agtcgaggcacacaacgacgagccacacaagcccaaaaaatctgacggcacagcaggtacaacc	6226
Db	6427	agtcgaggcacacaacgagccacacaagcccaaaaaatctgacggcacagcaggtacaacc	6486
Qy	6227	accacagcaggtgcaacccggtcacgtttaaaggcttctggacaaaacggcggttggtcg	6286
Db	6487	accacagcaggtgcaacccggtcacgtttaaaggcttctggacaaaacggcggttggtcg	6546
Qy	6287	gtctccgtgggtgcctccatggtgtgaacgcgcatccaaaaatgtggcagcaggtgaggtc	6346
Db	6547	gtctccgtgggtgcctccatggtgtgaacgcgcatccaaaaatgtggcagcaggtgaggtc	6606
Qy	6347	agtgccacagcagccgatacgggtcgaatgtgacagttgttacaaggccaccccaagcatt	6406
Db	6607	agtgccacagcagccgatacgggtcgaatgtgacagttgttacaaggccaccccaagcatt	6666
Qy	6407	gccaacgcaaccaatgagcttgaccatcgatatccaccaaaacgaaaaataagggccaatgcga	6466
Db	6667	gccaacgcaaccaatgagcttgaccatcgatatccaccaaaacgaaaaataagggccaatgcga	6726
Qy	6467	gggatttcacagcgatggcgatggcggtccatgcccacagcctcatctccctggcagatcc	6526
Db	6727	gggatttcacagcgatggcgatggcggtccatgcccacagcctcatctccctggcagatcc	6786
Qy	6527	atggtttaccgggggtattggcccccacaacgggtcgaagggtggcagttgggagctgtcg	6586
Db	6787	atggtttaccgggggtattggcccccacaacgggtcgaagggtggcagttgggagctgtcg	6846
Qy	6587	aagctgtcggataatgataatgggtatttaaaatcaatggttcagccgatacccaaggc	6646
Db	6847	aagctgtcggataatgataatgggtatttaaaatcaatggttcagccgatacccaaggc	6906
Qy	6647	catgtaggggcgcgagttgtgtgcagtttttcacttt	6682
Db	6907	catgtaggggcgcgagttgtgtgcagtttttcacttt	6942

RESULT	10	
AAF59129		
ID	AAF59129	standard; DNA; 720 BP.
XX	XX	
AC	AAF59129;	
XX	XX	
DT	24-APR-2001	(first entry)
XX	XX	
DE	M. catarrhalis strain 4223	200Kda partial nucleotide sequence.
XX	XX	
KW	Moraxella catarrhalis strain 4223;	major outer membrane protein;
KW	200Kda outer membrane protein;	antibacterial; immunogenic; infection;
KW	otitis media; detection; ds.	
XX	XX	
OS	Moraxella catarrhalis.	
XX	XX	
PN	WO200107619-A1.	
XX	XX	
PD	01-FEB-2001.	
XX	XX	
PF	26-JUL-2000;	2000WO-CA00870.
XX	XX	









```
Db 2948 ttgtcaggcaacaacaaatagcgaaggtataaacaggcgttgacgagggtgttggttacacag 3007
      || || || || || || || || || || || || || || || || || || || || || ||
Qy 6680 ttttaa 6685
      || || || || || || || || || || || || || || || || || || || || || ||
Db 3008 tggtaa 3013

RESULT 13
AAA92493
ID AAA92493 standard; DNA; 3036 BP.
XX
AC AAA92493;
XX
DT 17-JAN-2001 (first entry)
XX
DE Haemophilus influenzae adhesin (Hia) gene from NTHi strain 33.
XX
KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
KW diagnosis; immunogenic; antigen; ds.
XX
OS Haemophilus influenzae.
XX
PN WO200055191-A2.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-CA00289.
XX
PR 16-MAR-1999; 99US-0268347.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
XX WPI; 2000-618897/59.
DR P-PSDB; AAB23854.
XX
XX Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating Hemophilus influenzae
PT infection -
XX
XX Claim 1; Fig 18; 275pp; English.
XX
XX The present sequence represents a Haemophilus influenzae adhesin (Hia)
CC gene from the non-typeable Haemophilus influenzae (NTHi) strain 33.
CC Hia genes and proteins have antiinflammatory, auditory and antibacterial
CC activities, and can be used in the production of a vaccine. An
CC immunogenic composition comprising an Hia gene, a polypeptide encoded
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
CC protection against disease caused by Haemophilus strains in a
CC susceptible host, preferably a human. An Hia protein is useful as an
CC antigen, in immunogenic preparations including vaccines, as a carrier
CC for other immunogens, and in the generation of diagnostic reagents. Hia
CC is useful for treating diseases caused by the infection of Haemophilus
CC influenzae such as meningitis, epiglottitis, septicaemia and otitis
CC media. Recombinant production of Hia favours high recovery of the
CC protein compared to the low recovery of native protein from Haemophilus
CC influenzae species. A truncated protein has a significantly higher
CC amount of recovery than a full-length protein.
XX
XX Sequence 3036 BP; 965 A; 579 C; 799 G; 693 T; 0 other;
SQ

Query Match 2.18; Score 146.4; DB 21; Length 3036;
Best Local Similarity 58.28; Pred. No. 1.6e-27;
Matches 283; Conservative 0; Mismatches 191; Indels 12; Gaps 1;

Qy 6200 aaatctgacgcagcaggtacacaccacacagcaggtgcaaccggtacggttaaacgc 6259
      ||||| || || || || || || || || || || || || || || || || || || || || || ||
Db 2555 aaatctgagtgcggcggttaacactacaactcgtgtgtacacgggtacggttaaacgc 2614
```

```
Qy 6260 ttgtctgcaaaaacgagcggttggtgcgtctcctcgtggtgctcaggtgctgaacgcgt 6319
      ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2615 ttgtccggtgcaacggcgcaagcgtgctgctcgtcgcgcaagcgcggaagaagcgt 2674
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 6320 atccaaaatgtgcagcagcaggtgagtcagtgccacacacacacacacacacacacac 6379
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2675 atccaaaacgttgccgagcggaattccgctactccacacacacacacacacacacac 2734
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 6380 cagttgtacaaagccacaaagcattgccaacgcacaaatgagcttgaccatcgatc 6439
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2735 cagttgtatgcgtggcaaaaggggtatacaaac-----ctgctggacaagt 2782
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 6440 caccaaaacgaaataagcccaatcgagggtattccatcagcagcagcagcagcagcagc 6499
      || || || || ||||| || || || || || || || || || || || || || || ||
Db 2783 aataaagtgggcaaacgctgcagatgcaggtacagcaagtgcatcagcgttcacagtt 2842
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 6500 ccacaagcctacattcctgcagatccatggttacgggggtattgccacacacacacacac 6559
      ||||| ||||| || || || || || || || || || || || || || || || || || ||
Db 2843 ccacaagcctatgcaggttaataatcgaatggtttctattgcgggaagtattatcaagt 2902
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 6560 caagtgcggtgagcagtggaactgcgaagcgtgcgataatggtcaatgggtattttaa 6619
      ||||| ||||| || || || || || || || || || || || || || || || || || ||
Db 2903 caagtggttagctatcggggtatcaagaattccgataatcgcaagtgattattcgc 2962
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 6620 atcaatggttcagccgatacccaagccatgtaggggcgagcagtggtgaggtttcac 6679
      || || || || ||||| || || || || || || || || || || || || || || ||
Db 2963 ttgtcaggcacaaccaatagccaaggtataaacaggcgttcgacgaggtgtggttacc 3022
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 6680 ttttaa 6685
      || || || || || || || || || || || || || || || || || || || || || ||
Db 3023 tggtaa 3028

RESULT 14
AAA92495
ID AAA92495 standard; DNA; 3354 BP.
XX
AC AAA92495;
XX
DT 17-JAN-2001 (first entry)
XX
DE Haemophilus influenzae adhesin (Hia) gene from NTHi strain K9.
XX
KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
KW diagnosis; immunogenic; antigen; ds.
XX
OS Haemophilus influenzae.
XX
PN WO200055191-A2.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-CA00289.
XX
PR 16-MAR-1999; 99US-0268347.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
XX WPI; 2000-618897/59.
DR P-PSDB; AAB23858.
XX
XX Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating Hemophilus influenzae
PT infection -
XX
XX Claim 1; Fig 22; 275pp; English.
XX
XX The present sequence represents a Haemophilus influenzae adhesin (Hia)
```

KW	diagnosis; immunogenic; antigen; ds.
XX	Haemophilus Influenzae.
OS	
XX	
PN	W0200055191-A2.
XX	
XX	21-SEP-2000.
PD	
PF	16-MAR-2000; 2000WO-CA00289.
XX	
XX	16-MAR-1999; 990US-0266347.
PR	
PA	(CONN-) CONNAUGHT LAB LTD.
XX	
XX	Loosmore SM, Yang Y, Klein MH;
PI	
XX	WPI: 2000-618897/59.
DR	P-PSDB: AAB23859.
DR	

Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for use as antigens and vaccines and for treating Hemophilus influenzae infection -  
 Claim 1: Fig 23: 275pp: English.

Claim 1; Fig 23; 275pp; English.

The present sequence represents a Haemophilus influenzae adhesin (Hia) gene from the non-typeable Haemophilus influenzae (NTHi) strain K22. Hia genes and proteins have anti-inflammatory, auditory and antibacterial activities, and can be used in the production of a vaccine. An immunogenic composition comprising an Hia gene, a polypeptide encoded by an Hia gene, or a recombinant Hia polypeptide is useful for inducing protection against disease caused by Haemophilus strains in a susceptible host, preferably a human. An Hia protein is useful as an antigen, in immunogenic preparations including vaccines, as a carrier for other immunogens, and in the generation of diagnostic reagents. Hia is useful for treating diseases caused by the infection of Haemophilus influenzae such as meningitis, epiglottitis, septicaemia and otitis media. Recombinant production of Hia favours high recovery of the protein compared to the low recovery of native protein from Haemophilus influenzae species. A truncated protein has a significantly higher amount of recovery than a full-length protein.

Sequence 3342 BP; 1072 A; 641 C; 865 G; 764 T; 0 other;

```
Query Match          1.9%; Score 131.2; DB 21; Length 3342;
Best Local Similarity 57.4%; Pred. NO. 1.5e-23;
Matches 279; Conservative 0; Mismatches 198; Indels 9; Gaps 2
```

y	6380	cgattacaaagccaccceaaagcattgccaacgcaccaaatgagcttgaccatcgatc	6439
b	3032	cagttgatgcttgccaaggggtaaacaaatctgc---tgacaagtgaataaagtg	3088
y	6440	caccaaacgaaataaggccaatcagggtattcatccagcgtgcgatggcgctccatg	6499
b	3089	ggcaacgctgcagatgcaggtacagcaagtcgattagcagcttcacagttacc-----a	3142
y	6500	cccaagcctcacattcctgcagatccatggttaccgggggtattgccacccacaacgggt	6559
	3143	cccaacccctctatccagataatcaatgaattctctatttcggagaagtaagtattacaaggt	3202

0000 cccacggcaccacgagctatcttctattgcggggaagtgttatcaaggct 3202  
y ||||| | | | | | | | | | |  
b ccacaaacctctatgccacgtaaaatcaatggtttctattgcggggaagtgttatcaaggct 3202

Search completed: September 12, 2001, 13:21:22  
Job time: 2504 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 12:39:39 ; Search time 28653.2 Seconds  
(without alignments)  
3595.553 Million cell updates/sec

Title: US-09-361-619-5

Perfect score: 6972

Sequence: 1 ccattgatattggcagggtgt.....ccgtgacctgcaggtcgac 6972

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 28310096

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Pending\_Patents\_NA\_Main.\*
- 1: /cgn2\_6/ptodata/2/pna/PCTUS\_COMB.seq.\*
  - 2: /cgn2\_6/ptodata/2/pna/US06\_COMB.seq.\*
  - 3: /cgn2\_6/ptodata/2/pna/US07\_COMB.seq.\*
  - 4: /cgn2\_6/ptodata/2/pna/US080\_COMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/pna/US081\_COMB.seq.\*
  - 6: /cgn2\_6/ptodata/2/pna/US082\_COMB.seq.\*
  - 7: /cgn2\_6/ptodata/2/pna/US083\_COMB.seq.\*
  - 8: /cgn2\_6/ptodata/2/pna/US084\_COMB.seq.\*
  - 9: /cgn2\_6/ptodata/2/pna/US085\_COMB.seq.\*
  - 10: /cgn2\_6/ptodata/2/pna/US086\_COMB.seq.\*
  - 11: /cgn2\_6/ptodata/2/pna/US087\_COMB.seq.\*
  - 12: /cgn2\_6/ptodata/2/pna/US088\_COMB.seq.\*
  - 13: /cgn2\_6/ptodata/2/pna/US089\_COMB.seq.\*
  - 14: /cgn2\_6/ptodata/2/pna/US090\_COMB.seq.\*
  - 15: /cgn2\_6/ptodata/2/pna/US091\_COMB.seq.\*
  - 16: /cgn2\_6/ptodata/2/pna/US092\_COMB.seq.\*
  - 17: /cgn2\_6/ptodata/2/pna/US093\_COMB.seq.\*
  - 18: /cgn2\_6/ptodata/2/pna/US094\_COMB.seq.\*
  - 19: /cgn2\_6/ptodata/2/pna/US095A\_COMB.seq.\*
  - 20: /cgn2\_6/ptodata/2/pna/US095B\_COMB.seq.\*
  - 21: /cgn2\_6/ptodata/2/pna/US095C\_COMB.seq.\*
  - 22: /cgn2\_6/ptodata/2/pna/US095D\_COMB.seq.\*
  - 23: /cgn2\_6/ptodata/2/pna/US096A\_COMB.seq.\*
  - 24: /cgn2\_6/ptodata/2/pna/US096B\_COMB.seq.\*
  - 25: /cgn2\_6/ptodata/2/pna/US096C\_COMB.seq.\*
  - 26: /cgn2\_6/ptodata/2/pna/US096D\_COMB.seq.\*
  - 27: /cgn2\_6/ptodata/2/pna/US096E\_COMB.seq.\*
  - 28: /cgn2\_6/ptodata/2/pna/US097A\_COMB.seq.\*
  - 29: /cgn2\_6/ptodata/2/pna/US097B\_COMB.seq.\*
  - 30: /cgn2\_6/ptodata/2/pna/US097C\_COMB.seq.\*
  - 31: /cgn2\_6/ptodata/2/pna/US098\_COMB.seq.\*
  - 32: /cgn2\_6/ptodata/2/pna/US6000\_COMB.seq.\*
  - 33: /cgn2\_6/ptodata/2/pna/US6001\_COMB.seq.\*
  - 34: /cgn2\_6/ptodata/2/pna/US6002\_COMB.seq.\*
  - 35: /cgn2\_6/ptodata/2/pna/US6003\_COMB.seq.\*
  - 36: /cgn2\_6/ptodata/2/pna/US6004\_COMB.seq.\*
  - 37: /cgn2\_6/ptodata/2/pna/US6005\_COMB.seq.\*
  - 38: /cgn2\_6/ptodata/2/pna/US6006\_COMB.seq.\*
  - 39: /cgn2\_6/ptodata/2/pna/US6007\_COMB.seq.\*
  - 40: /cgn2\_6/ptodata/2/pna/US6008\_COMB.seq.\*
  - 41: /cgn2\_6/ptodata/2/pna/US6009\_COMB.seq.\*
  - 42: /cgn2\_6/ptodata/2/pna/US6010\_COMB.seq.\*
  - 43: /cgn2\_6/ptodata/2/pna/US6011\_COMB.seq.\*

- 44: /cgn2\_6/ptodata/2/pna/US6012\_COMB.seq.\*
- 45: /cgn2\_6/ptodata/2/pna/US6013\_COMB.seq.\*
- 46: /cgn2\_6/ptodata/2/pna/US6014\_COMB.seq.\*
- 47: /cgn2\_6/ptodata/2/pna/US6015\_COMB.seq.\*
- 48: /cgn2\_6/ptodata/2/pna/US6016\_COMB.seq.\*
- 49: /cgn2\_6/ptodata/2/pna/US6017\_COMB.seq.\*
- 50: /cgn2\_6/ptodata/2/pna/US6018\_COMB.seq.\*
- 51: /cgn2\_6/ptodata/2/pna/US6019\_COMB.seq.\*
- 52: /cgn2\_6/ptodata/2/pna/US6020\_COMB.seq.\*
- 53: /cgn2\_6/ptodata/2/pna/US6021\_COMB.seq.\*
- 54: /cgn2\_6/ptodata/2/pna/US6022\_COMB.seq.\*
- 55: /cgn2\_6/ptodata/2/pna/US6023\_COMB.seq.\*
- 56: /cgn2\_6/ptodata/2/pna/US6024\_COMB.seq.\*
- 57: /cgn2\_6/ptodata/2/pna/US6025\_COMB.seq.\*
- 58: /cgn2\_6/ptodata/2/pna/US6026\_COMB.seq.\*
- 59: /cgn2\_6/ptodata/2/pna/US6027\_COMB.seq.\*
- 60: /cgn2\_6/ptodata/2/pna/US6028\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	6972	100.0	6972	17	US-09-361-619-5
2	6961	99.8	6973	8	US-08-483-855-1
3	6961	99.8	6973	10	US-08-621-944-1
4	6961	99.8	6973	10	US-08-621-944A-1
5	6961	99.8	6973	13	US-08-945-567-1
6	6961	99.8	6973	13	US-08-945-567A-1
7	6961	99.8	6973	13	US-08-945-567C-1
8	6961	99.8	6973	13	US-08-945-567D-1
9	6961	99.8	6973	17	US-09-361-619-1
10	6939	99.5	6975	8	US-08-431-718-1
11	6939	99.5	6975	8	US-08-431-718A-1
12	6939	99.5	6975	8	US-08-431-718C-1
13	6258	89.8	6259	17	US-09-361-619-12
14	6141	88.1	6141	17	US-09-361-619-6
15	5976	85.7	5976	10	US-08-621-944-2
16	5976	85.7	5976	10	US-08-621-944A-2
17	5976	85.7	5976	13	US-08-945-567-2
18	5976	85.7	5976	13	US-08-945-567A-2
19	5976	85.7	5976	13	US-08-945-567C-2
20	5976	85.7	5976	13	US-08-945-567D-2
21	5976	85.7	5976	17	US-09-361-619-2
22	4671.2	67.0	62909	22	US-09-596-002-32
23	4671.2	67.0	62909	46	US-60-140-121-32
24	4655.8	66.8	15958	38	US-60-068-139-381
25	4635.8	66.8	17368	40	US-60-082-116-290
26	4396.8	63.1	9542	31	US-09-813-214-9
27	4101.6	58.8	6159	17	US-09-361-619-8
28	2590	37.1	6942	17	US-09-361-619-10
29	2007.2	28.8	6429	21	US-09-540-236-1539
30	2007.2	28.8	6429	44	US-60-128-476-1992
31	220.8	3.2	1005	21	US-09-540-236-1266
32	220.8	3.2	1005	44	US-60-128-476-2460
33	148	2.1	3030	16	US-09-268-347-29
34	146.4	2.1	3036	16	US-09-268-347-23
35	121.4	1.7	123	44	US-60-128-476-1993
36	85.4	1.2	7486	14	US-09-077-098-5
37	85.4	1.2	7486	14	US-09-077-098A-5
38	84.4	1.2	6091	18	US-09-453-702B-200
39	83.8	1.2	2079	16	US-09-268-347-25
40	83.8	1.2	8930	14	US-09-077-098A-1
41	83.8	1.2	8930	14	US-09-077-098A-1
42	82.8	1.2	3294	16	US-09-268-347-43
43	78.4	1.1	1812	16	US-09-268-347-37
44	78.2	1.1	3300	16	US-09-268-347-31
45	78.2	1.1	7253	16	US-09-268-347-35





QY 1741 aaaagaggctgataataagtgctgaaagttaaacttgctaaaaactttaacaatctttac 1800  
Db 1741 aaaagaggctgataataagtgctgaaagttaaacttgctaaaaactttaacaatctttac 1800  
QY 1801 tgaggtaatacaactacattaaatgcacaaacacacaggttaagtagtagtagtagtag 1860  
Db 1801 tgaggtaatacaactacattaaatgcacaaacacacaggttaagtagtagtagtagtag 1860  
QY 1861 tactacagtgtaattattagtgtagtagtttaacctttaccagcccaatacagcgagtca 1920  
Db 1861 tactacagtgtaattattagtgtagtagtttaacctttaccagcccaatacagcgagtca 1920  
QY 1921 aagcacaaagcaaaacccgtctatgctgtaataatggggtgaagtttactaataatgcagaaac 1980  
Db 1921 aagcacaaagcaaaacccgtctatgctgtaataatggggtgaagtttactaataatgcagaaac 1980  
QY 1981 aacagcagcaatcgccactactcgtattaccagagataaaaattggcttgcgcagatgg 2040  
Db 1981 aacagcagcaatcgccactactcgtattaccagagataaaaattggcttgcgcagatgg 2040  
QY 2041 tgatgttgatgaaaaaacaagcaccattatttggataaaaaaacaacttaaaagggttagtgt 2100  
Db 2041 tgatgttgatgaaaaaacaagcaccattatttggataaaaaaacaacttaaaagggttagtgt 2100  
QY 2101 tgcaattaccatagacaatggcatttgatgcaggttaataaaaaagatcagtaattcgcaa 2160  
Db 2101 tgcaattaccatagacaatggcatttgatgcaggttaataaaaaagatcagtaattcgcaa 2160  
QY 2161 aggtagcagtgctaacgatcggtgtaccatcgaaacagctcaaaagccgcaagcctacttt 2220  
Db 2161 aggtagcagtgctaacgatcggtgtaccatcgaaacagctcaaaagccgcaagcctacttt 2220  
QY 2221 aaacgaggcgctggcctacagtgctcacacctactgaaataatcagttgatgtaagatgg 2280  
Db 2221 aaacgaggcgctggcctacagtgctcacacctactgaaataatcagttgatgtaagatgg 2280  
QY 2281 caatgttacgcccccacttacacattggcgtgaaacaccagcagcttaacagtgatgg 2340  
Db 2281 caatgttacgcccccacttacacattggcgtgaaacaccagcagcttaacagtgatgg 2340  
QY 2341 cactagtgataaatttagtggtaaggttagtgggtacgaaacaataagcttagttacgcgcga 2400  
Db 2341 cactagtgataaatttagtggtaaggttagtgggtacgaaacaataagcttagttacgcgcga 2400  
QY 2401 acatttggcaagctatctaaatgaagtcgaatcgacacggtgacagtgctctacaaagctt 2460  
Db 2401 acatttggcaagctatctaaatgaagtcgaatcgacacggtgacagtgctctacaaagctt 2460  
QY 2461 taccgttaaaagaagaacgatgatgacgcgaacagctatacccggtggttaaaagatacgac 2520  
Db 2461 taccgttaaaagaagaacgatgatgacgcgaacagctatacccggtggttaaaagatacgac 2520  
QY 2521 aaaaaatgcggcgagtcagcatcttaaaactcaaaaggttaaaacggtctaaacggttgc 2580  
Db 2521 aaaaaatgcggcgagtcagcatcttaaaactcaaaaggttaaaacggtctaaacggttgc 2580  
QY 2581 taccataaaagatggttacggttacctttggcttagccaagatagcggctgcaccattgg 2640  
Db 2581 taccataaaagatggttacggttacctttggcttagccaagatagcggctgcaccattgg 2640  
QY 2641 caaaagcaccctaaacaacgatgcttgactgttaaaagatacccaagcaaaaatcccaagt 2700  
Db 2641 caaaagcaccctaaacaacgatgcttgactgttaaaagatacccaagcaaaaatcccaagt 2700  
QY 2701 cgggtgtaattggccttaaaattactaatgtgaaatgggtagtaatccaggtactggcattgc 2760  
Db 2701 cgggtgtaattggccttaaaattactaatgtgaaatgggtagtaatccaggtactggcattgc 2760  
QY 2761 aaataccgcgtcgattaccagagataaaattggcttggcttctggttctgatgctgcagttga 2820  
Db 2761 aaataccgcgtcgattaccagagataaaattggcttggcttctggttctgatgctgcagttga 2820  
QY 2821 taaaaaacaaccttatcttgatcaagacaagctacaagttggcgaatgttaagattaccaa 2880

Db 2821 taaaaaacaaccttatcttgatcaagacaagctacaagttggcaatgttaagattaccaa 2880  
QY 2881 cactggcattaacgcaggtggttaaagccatcacagggtgtccccacacgtcctagcat 2940  
Db 2881 cactggcattaacgcaggtggttaaagccatcacagggtgtccccacacgtcctagcat 2940  
QY 2941 tgcgcgtacaaagtagcgcgaacatagaaactgggcaatacaatcccaagcaaaagacaatc 3000  
Db 2941 tgcgcgtacaaagtagcgcgaacatagaaactgggcaatacaatcccaagcaaaagacaatc 3000  
QY 3001 caacgctgcagcatttaattgatatattaaatcacaggcttttaacctaaaaataataacaa 3060  
Db 3001 caacgctgcagcatttaattgatatattaaatcacaggcttttaacctaaaaataataacaa 3060  
QY 3061 cccattgacttctgctccacttatgacattgttgacttgcacttccaatggcaatgcccaccac 3120  
Db 3061 cccattgacttctgctccacttatgacattgttgacttgcacttccaatggcaatgcccaccac 3120  
QY 3121 cgcacagtagtaccctgataccgctaaacaaacacagtagtaagtggtatagtgtaattgt 3180  
Db 3121 cgcacagtagtaccctgataccgctaaacaaacacagtagtaagtggtatagtgtaattgt 3180  
QY 3181 ggatgatacaaccattcaatcaacaggcactgatgacaataaaaaacttggcgtcaaaac 3240  
Db 3181 ggatgatacaaccattcaatcaacaggcactgatgacaataaaaaacttggcgtcaaaac 3240  
QY 3241 caccaaactgaacaaaaacaagtcttaattgtaatacagacaactaacttaatttaactc 3300  
Db 3241 caccaaactgaacaaaaacaagtcttaattgtaatacagacaactaacttaatttaactc 3300  
QY 3301 tagtgagaagtgccttggtaacgcgaagacatacgccgaaaaatctaaacacccctagc 3360  
Db 3301 tagtgagaagtgccttggtaacgcgaagacatacgccgaaaaatctaaacacccctagc 3360  
QY 3361 caaggaattcacaccacccaaaggcacagcagacacccctacataaaactttaccggttaa 3420  
Db 3361 caaggaattcacaccacccaaaggcacagcagacacccctacataaaactttaccggttaa 3420  
QY 3421 aaaggtagatgaaaaataatgctgtagcgcgaacccatcacccgtgggtcgaagaa 3480  
Db 3421 aaaggtagatgaaaaataatgctgtagcgcgaacccatcacccgtgggtcgaagaa 3480  
QY 3481 cgcaataatcaagtcaaacacccataacactcaaaaggtgaaacggtcttaattataaac 3540  
Db 3481 cgcaataatcaagtcaaacacccataacactcaaaaggtgaaacggtcttaattataaac 3540  
QY 3541 cgcaaaaaatggtagcgtttacctttggcatttaacacacacagcgtcttaaaagccgcga 3600  
Db 3541 cgcaaaaaatggtagcgtttacctttggcatttaacacacacagcgtcttaaaagccgcga 3600  
QY 3601 aagcacccctaaacgagcgtggcttgccttatataaaacccctggtagcgaacaaatcca 3660  
Db 3601 aagcacccctaaacgagcgtggcttgccttatataaaacccctggtagcgaacaaatcca 3660  
QY 3661 agtcggtgctgatggcgtgaagtttgcgaaggttaataatgggtttaggtgctgg 3720  
Db 3661 agtcggtgctgatggcgtgaagtttgcgaaggttaataatgggtttaggtgctgg 3720  
QY 3721 cattgtaggcacaactcgcttaccagagatgaaattggcttactgggaactaatggctc 3780  
Db 3721 cattgtaggcacaactcgcttaccagagatgaaattggcttactgggaactaatggctc 3780  
QY 3781 acttgataaaagcaaaacccacccctaaagcaaaagcggcattcaacgaggtggttaaaagat 3840  
Db 3781 acttgataaaagcaaaacccacccctaaagcaaaagcggcattcaacgaggtggttaaaagat 3840  
QY 3841 taccacaattcaatcaggtgagatttgcgaacacagcagcatgatcgtgcagggcgcaa 3900  
Db 3841 taccacaattcaatcaggtgagatttgcgaacacagcagcatgatcgtgcagggcgcaa 3900  
QY 3901 gatttatgatttataaaacccgaacttgaaaaacaaactcagcagtagctgcacaaacagcaca 3960



Qy	6121	cgtagccgaagaagtaactcgggttgcccttaggttcaaaactctgcacatcagtcgagggcacaca	6180
Db	6121	cgtagccgaagaagtaactcgggttgcccttaggttcaaaactctgcacatcagtcgagggcacaca	6180
Qy	6181	cgagggcacacaagccaaaaatctgcgcgcacagcaggttacaaccaccacagcaggtgc	6240
Db	6181	cgagggcacacaagccaaaaatctgcgcgcacagcaggttacaaccaccacagcaggtgc	6240
Qy	6241	aaccggtacggttaaaaggcttctgtggacaaaacggcggttggctgcgtctccgttgggtgc	6300
Db	6241	aaccggtacggttaaaaggcttctgtggacaaaacggcggttggctgcgtctccgttgggtgc	6300
Qy	6301	ctcagtgctgaacgccgtatccaaaatgtggcagcaggtgagtcagtcagtcaccaccagcac	6360
Db	6301	ctcagtgctgaacgccgtatccaaaatgtggcagcaggtgagtcagtcagtcaccaccagcac	6360
Qy	6361	cgatcggtccaatggtgagcagttgtacaaggccaccaaaagcattgccaacgcaaccaa	6420
Db	6361	cgatcggtccaatggtgagcagttgtacaaggccaccaaaagcattgccaacgcaaccaa	6420
Qy	6421	tgaagttgacctcgatatccaccaaaacgaaaaaagaagccaatgcagggatttcacatcagc	6480
Db	6421	tgaagttgacctcgatatccaccaaaacgaaaaaagaagccaatgcagggatttcacatcagc	6480
Qy	6481	gatggcgatggcggtccatgctccacaagacctacattccttgccagatcccatggttacccgggg	6540
Db	6481	gatggcgatggcggtccatgctccacaagacctacattccttgccagatcccatggttacccgggg	6540
Qy	6541	tattgcaccccaacacggtccaaggtgcgggtggcgagtcgtcgaaagctgtcgagataa	6600
Db	6541	tattgcaccccaacacggtccaaggtgcgggtggcgagtcgtcgaaagctgtcgagataa	6600
Qy	6601	tggtcgaatgggtatttaaaatcaatgggttcagcgcgataccacaagggccatgtaggggcgcg	6660
Db	6601	tggtcgaatgggtatttaaaatcaatgggttcagcgcgataccacaagggccatgtaggggcgcg	6660
Qy	6661	agttggtcgaggttttcaacttttaagccataaatacgcaagatttacttaaaatcaatc	6720
Db	6661	agttggtcgaggttttcaacttttaagccataaatacgcaagatttacttaaaatcaatc	6720
Qy	6721	tcaccatggttatataaaacagcatcagcatcagtcattactgatgctgatgttttt	6780
Db	6721	tcaccatggttatataaaacagcatcagcatcagtcattactgatgctgatgttttt	6780
Qy	6781	atacattaaaccatttttaacgctcaagtgattctcttccacatgacccaatcgccattg	6840
Db	6781	atacattaaaccatttttaacgctcaagtgattctcttccacatgacccaatcgccattg	6840
Qy	6841	atacattggttaaaacttattgagttaaattttatcaatgtagtgttagatatggtttaaaatt	6900
Db	6841	atacattggttaaaacttattgagttaaattttatcaatgtagtgttagatatggtttaaaatt	6900
Qy	6901	gtgccattgacaaaataatgaccgatttatccgaaaaatttctgatgatgatcgcttgac	6960
Db	6901	gtgccattgacaaaataatgaccgatttatccgaaaaatttctgatgatgatcgcttgac	6960
Qy	6961	ctgcaggtcgac	6972
Db	6961	ctgcaggtcgac	6972

RESULT 2

```

US-08-483-855-1
; Sequence 1, Application US/08483855
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARNES, Robin E.
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

```

Db 481 CCCTAATTATTCAATCAATGCTATGTGACGATGATATCATTTTAAAGGTAAACCAC 540  
Qy 541 catgaatcacatctataaagtcacttttaacaaagccacaggcacatttatggcagtgcc 600  
Db 541 CATGAATACATCATATAAAGTCATCTTTAACAAAGCCACAGGCACATTTATGGCAGTGGC 600  
Qy 601 agagtgcgcaaatcccaagcac-gggggggggtagctgtgtacagggcgaagtgggca 659  
Db 601 AGAGTACGCCAAATCCACACACACGGGGGGGTAGCTGTCTACAGGGCAAGTTGGCA 660  
Qy 660 gtgtatgcaactgtagottttgcgcgtatttgcgcgtctgcgtctgcgtgcgtgcgtgc 719  
Db 661 GTGTATGCACTGTGAGCTTTGCCGTTATTTGCCGCTGCTGCTGCTGCTGCTGCTGCTG 720  
Qy 720 cgtcagtgagcagtgcttgcctcaaaaaaagatatcccaaatatcgcaattgggcaac 779  
Db 721 CGCTACGTGGCAGTGTATGCTCAAAAAAAGATACCAAAATATCGCAATTTGGTGAAC 780  
Qy 780 aaaaaccagcaagacgctcaggcactgcgaagcggagcgggtagcgcagccattgctattg 839  
Db 781 AAAACCAAGCAAGGCTCAGGCACATGCCAAGCGGACGGTGTATGCTATTTG 840  
Qy 840 gtgaataatgcacagcagcagggcggtcaagccatgcgccatcggttagtagtaataaaactg 899  
Db 841 GTGAAAATGCTAAACGACAGGGGGGTCAAGCCATCGCCATCGGTAGTAGTAATAAAACTG 900  
Qy 900 tcaatgaagcagtttgataaagtagtaccgatgctacgggtcgaagagtcacatcgcca 959  
Db 901 TCAATGGAACGAGTTTGGATAGATAGGTACCGATGCTACGGGTCAAGAGTCCATCGCCA 960  
Qy 960 tcggtggtgataaaagcctagtggtgtagtcctcgatggccatcggttagtgatgaactac 1019  
Db 961 TCGGTGTGTATTAAGGCTAGTGTGATGCTGATGCTCGATGCTCGGTAGTATGATCTTAC 1020  
Qy 1020 atttggtgtagcagtaggtaacttaaacatccgaagtagtctgattaaagcagctta 1079  
Db 1021 ATTTGCTTGATCAGCATGGTAATCCTTAACATCCGAAGGTACTCTGATTAAAGATCTTA 1080  
Qy 1080 ttaacggcctagcagtagtataaagaatacgaagctcaaaaggataatgatgtaaaata 1139  
Db 1081 TTAACGGCCATGCAGTATTAAGAAATACGAAGCTCAAAAGGATAATGATGTAATAATA 1140  
Qy 1140 gacgcacaaacgcaagcggacacccagtagtgcagtgaggagccatgcatatgcacag 1199  
Db 1141 GACGCACAACCGCAAGCGGACACGCCAGTCTGCAGTGGGAGCCATGCTATGTCACAGG 1200  
Qy 1200 gtcatttttccaaacgctttgtgtacagggcaacagctaaaaagtgccctattccttggcag 1259  
Db 1201 GTCATTTTCCACGCTTTGCTACACGGGCAACAGCTAAAAAGTGCCTATTCTCTTGGCAG 1260  
Qy 1260 ttgggtcttgcgcgcacagcggagggccaatctacaatcgctattggtttctgtagcaacat 1319  
Db 1261 TGGGTCTTTGCCGCCACAGCCGAGGGCCCAATCTACAAATCGCTATTGTTCTGATGCAACAT 1320  
Qy 1320 ctatgctgttgaggagcagtagcccttgggtcaggtactcgtgcagctacagagggcagta 1379  
Db 1321 CTAGCTCTGTTGGGACCGATACCCCTTGGTGCAGGTACTCTGCTCAGCTACAGGGCAGTA 1380  
Qy 1380 ttgcccctaggttcaaggttctgtgtcactcagagtgtataataatctagacccggcctata 1439  
Db 1381 TTGCCCTTAGTCAAGGTTCTGTGCTACTCAGAGTGATAATAATCTAGACGGGCTATA 1440  
Qy 1440 caccaaaataccaggcactagaccccaagtttcaagccacaataatacgaagcgagtc 1499  
Db 1441 CACCAAAATACCCAGGCATAGACCCCAAGTTTCAAGCCACCATAATACGAAGCGGGTGC 1500  
Qy 1500 cacttccattggttagtaactctataacagtaaaatcataatcaatgctgggtcaggtgtta 1559  
Db 1501 CACTTTCCATTGGTAGTAACCTATCAACAGTAAAAATCATCAATGTCGGTGCAGGTGTTA 1560  
Qy 1560 ataaacccagatgcgtcgaattgtgcacagcgtagaaagcgttggtgaagtcggtcgaagagc 1619  
Db 1561 ATAAAACCGATGCGGTCAATGTGGCACAGCTAGAAAGCGGTGGAAGTGGGCTAAGGAGC 1620

Qy 1620 gtagaattacttttcaggggtgatgataacagctactgacgtataaaataggtttggataata 1679  
Db 1621 GTAGAAATTTACTTTTCAGGGTGATGATAACAGTACTGACTAAAAATAGTGTGGATAATA 1680  
Qy 1680 ctttaactattaaaagtggtgcagagacacacgcatataacgagataataatcgggtgag 1739  
Db 1681 CTTTAACTATTAAAGGTGTGCAGAGACCAACGCATTAAACGATTAATAATATCGGTGTGG 1740  
Qy 1740 taaaagagcgtgataatagtggtctgaaagttaaacttgctaaaaactttaaacaacttta 1799  
Db 1741 TAAAGAGGCTGATTAATAGTGTCTGAAAGTTAACTTGCATAAACTTTAAACAATCTTA 1800  
Qy 1800 ctgaggtgaatacaactacattaaatgccaacacacagtttaaggtaggtagtagtagta 1859  
Db 1801 CTGAGGTGAATACAACACTACATTTAAATGCCACAAACACAGTTAAGGTAGGTAGTAGTA 1860  
Qy 1860 gtactacagctgaattattagtgatagtttaacctttaccacgcccatacagcagcagtc 1919  
Db 1861 GTACTACAGCTGAATTTTGTAGTGATAGTTTAACTTTACCCAGCCCATATACAGCAGTC 1920  
Qy 1920 aaagcacaagcaaaacgcgtctatggttaattggtggtgaagtgttactaataatgcagaaa 1979  
Db 1921 AAAGCACAGCAAAACCCGCTATGCGTTAATGGGTGAAGTTTACTAATAATGCAGAAA 1980  
Qy 1980 caacagcagcaatcggcactactcgtattaccagagataaaattggctttgctcagagatg 2039  
Db 1981 CAACAGCAGCAATCGGCACACTACTCGTATTTACCAGAGATAAAATTTGCTTGTCTCAGATG 2040  
Qy 2040 gtgattgataaaaaaacagcaccatatttgataaaaaaacacttaagtgggtgagtg 2099  
Db 2041 GTGATGTTGATGAAAAAACAAGCACCATATTTGGATAAAAAACAACNTTAAAGTGGGTAGTG 2100  
Qy 2100 ttgcaattaccatagacaaatggcattgtagcagggttaataaaagatcagtaactcgtgcca 2159  
Db 2101 TTGCAATTACCATAGACAATGGCATTGATGACAGTAAATAAAAGATCAGTAATCTTGCCA 2160  
Qy 2160 aagtgtagcagtgctaaagcagtggtttaccatgcgaacagctcaaaagcccaagcctactt 2219  
Db 2161 AAGGTAGCAGTGTGTAACGATGCGGTTTACCATGCAACAGCTCAAAAGCCGCAAGCCTACTT 2220  
Qy 2220 taaacagcagcgtcgtgcatcagtgctcacacctactgaaataatcagttgagtgcataagcgtg 2279  
Db 2221 TAAACGACGCGCTGCGCATCAGTGTACACCTACTCAAAATATCAGTTGATGCTAAGAGTG 2280  
Qy 2280 gcaattgtaccgccccaaacttaacaacattggcgtgaaacccacagcagcttaacagtgatg 2339  
Db 2281 GCAATGTTACCGCCCAACTTACAACATTGGCGGTGAAACCCACCGAGCTTAACAGTGTATG 2340  
Qy 2340 gcactagtgataaaatttagtttaagggtagtggtacgaacaatagcttagttaccgccc 2399  
Db 2341 GCACGTAGTATAAATTTAGTGTTAAGGTAGTGTGACGAACAATAGCTTAGTTACCGCCG 2400  
Qy 2400 aacatttggcaagctatctaaatgaagtcgaatgcgaacggtgcagctgctctcaaaagct 2459  
Db 2401 AACATTGGCAAGCTATCTAAATGAAGTCAATCGAAGGCTGACAGTGTCTTACAAGCT 2460  
Qy 2460 ttaccgttaagaagaagacgatgacgcgaacgctatcacccgtgggtcaasagatacga 2519  
Db 2461 TTACCGTTAAAGAAAGAAAGACGATGATGACGCCAACGCTATCACCGTGGCTTAAGATACGA 2520  
Qy 2520 caaaaaatgcgcgcagtcagcatcttaaaactcaaaagctaaaaacgggtctaacgggtg 2579  
Db 2521 CAAAAATGCCGCGCAGTCAGCATCTTAAACTCAAAAGGTAAAAACGGTCTAACGGTTG 2580  
Qy 2580 ctacaaaaaagatggtaacggttaccttttggttagccaagatagcggtctacacattg 2639  
Db 2581 CTACCAAAAAAAGATGGTACGGTTTACCTTTGGGCTTACCCAAAGATAGCGGTCTGACCAATTG 2640  
Qy 2640 gcaaaagcacctcaaaacacagatggtgtgactgttaaagataccacgaacgaacaaatccaag 2699  
Db 2641 GCAAAACACCTTAACACAGATGGCTTGACTGTTAAAGATACCAACGAACAAATCCAAG 2700

QY 2700 tcggtgctaagtgcattaaattactaataatggaatggtagtaataaccaggctactggcattg 2759  
Db 2701 TCGGTGCTAAATGGCAATTAATTTACTAATGTAATGGTAGTAATCCAGGTACTGGCATTTG 2760  
QY 2760 caaataccgctcgcattaccagagataaaattggcttggctggttctctgatggtgcagttg 2819  
Db 2761 CAAATACCGCTCGCATTAACAGAGATAAAATGGCTTTGCTGGTTCYCATGGTGCAGTTG 2820  
QY 2820 atacaaacaaaccccttatcttgatcaagacaagctacaaagttggcaattgtaagattacca 2879  
Db 2821 ATACAAACAAACCTTATCTTGATCAAGACAAGCTACAAGTTGGCAATGTTAAGATTACCA 2880  
QY 2880 aactggcattaaagcaggtggttaagaccatcaagggctgtccccaacactgacctagca 2939  
Db 2881 ACACTTGGCATTAAAGCAGGTGGTAAGCCATCACAGGGCTGTCCCAACACTGGCTAGCA 2940  
QY 2940 ttgcgcatcaagtagccgcgaactagaaactgggcaatacaactccaagacaagaacaat 2999  
Db 2941 TTGGCGATCAAGTAGCCGCAACATAGAACTGGGCAATACAAATCCAGACAAAGACAAT 3000  
QY 3000 ccaacgctgcagcattaatgatataataacagggctttaacctaaacaaataataaaca 3059  
Db 3001 CCAACGCTGCCAGCATTAATGATATATAATACAGGCTTTAAACCTAAATAATAACA 3060  
QY 3060 acccattgaacttgcctccacttatgacattgttgactttggcaattggcaatggcaacca 3119  
Db 3061 ACCCCATTGACTTTGCTCCACTTATGACATGTGTGACTTTGGCAATGGCAATGCCACCA 3120  
QY 3120 cgcgcacagtaaccatgataccgctaaacaaacccagtaaaagtggtatgatgtgaatg 3179  
Db 3121 CGGCCACAGTAACCCATGATACCGGTACAAACCCAGTAAGTGGTATGATGTGAATG 3180  
QY 3180 tggatgatacaacacattcatctaacaggcactgatgacaataaaaaacttggcgtcaaaa 3239  
Db 3181 TGGATGATACAACCATTCATTAACAGGCACCTGATGACAATAAAAACTTGGCGTCAAAA 3240  
QY 3240 ccacaaactgaacaaacaaagtgctaattgtaatacagcaactaacttctaatgttaact 3299  
Db 3241 CCACCAACTGAACAAAACAAAGTGTGTAATGTTATACAGCAACTAACTTTAATGTTAACT 3300  
QY 3300 ctatgtagaagatgaccttgttaacgccaagacatcgccgaataataacacacccctag 3359  
Db 3301 CTAGTGATGAAGTGCCTTGTTAACGCCAAGACATCGCCGAAATCTAAACACCCCTAG 3360  
QY 3360 ccaaggaaatttcacaccaccaagagcagcagacacccgctcacaacccctttaccggtta 3419  
Db 3361 CCAAGAAATTCACACCACCAAGGACAGCAGACACCGCCCTACAAACCTTTACCGTTA 3420  
QY 3420 aaaaggtagatgaaaaataatgctgatgacgccaacgccaatcacccgtgggtcaaaaaga 3479  
Db 3421 AAAAGGTAGATGAAAATAATATGCTGATGACGCCAACGCCATCACCGTGGGTCAAAAAGA 3480  
QY 3480 acgcaataataaagtcacaccccttaacactcaaaagtgaaaacggtcttaataataaaa 3539  
Db 3481 ACGCAATAATCAAGTCAACACCCCTAACACTCAAAAGGTGAACACGGTCTTAAATATAAA 3540  
QY 3540 ccgacaaaaatgggtacggttacctttggcattaaacaccacagcggtcttaaaagccggca 3599  
Db 3541 CCGCAAAAAATGGTAGCGTTACCTTTGGCAATTAACACCACAAGCGGTCTTAAAGCCGGCA 3600  
QY 3600 aaagcacccctaaacgagcgtggctgtctataaaaaacccactggtagcgaaacaaatccc 3659  
Db 3601 AAAGCACCTTAACGACGGTGGCTGTCTATTAAAAACCCCACTGGTAGCGAACAATCC 3660  
QY 3660 aagtcggtgctgtagcgtggaatttgcgaaggttaataataatgggtgttaggtgctg 3719  
Db 3661 AAGTCGTGCTGATGGCTGAAGTTTGCAAGGTAAATAATAATGCTTTGTAGGTGCTG 3720  
QY 3720 gcatgtggcacactgcattaccagagatgaaattggctttactgggactaatggct 3779  
Db 3721 GCATTGATGGCAACACTCGCATTAACAGAGATGAAATTTGGCTTTACTGGGACTAATGGCT 3780  
QY 3780 cacttgataaagaacaaacccacccactaagcaaaagcgcattaaacgaggttggttaaaaaga 3839

Db 3781 CACTTGTATAAAGAGAAACCCACCTTAAGCAAAAGACGGCATTAACGACGGTGTAAAAAAGA 3840  
QY 3840 ttaccaacatcaatcaggtgagattgccccaaacagccatgatctgtgacagcgcca 3899  
Db 3841 TTACCAACATTCATCAATCAGTGAGATTTGCCAAAACAGCCATGATCTGTGACAGCGGCA 3900  
QY 3900 agattatgattataaaacccgaacttgaaacaaatcagcagctactgcaaaaacagcac 3959  
Db 3901 AGATTTATGATTTAAACACCGAACTTGAAACAAAAATCAGCAGTACTGCGCAAAACAGCAC 3960  
QY 3960 aaaaactattacacgaattctcagtagcagatgaacaaggtaataaactttacggttagta 4019  
Db 3961 AAAACTCATTAACCAAAATCTCAGTAGCAGATGAACAAGTAATAACTTTACGGTTAGTA 4020  
QY 4020 acccttactccagttatcacacctcaagacctctgagtcatcacccttgcagtgaaa 4079  
Db 4021 ACCCTTACTCCAGTTATGACACCTCAAGACCTCTGATGTCAATCACTTTGCGAGGTGAAA 4080  
QY 4080 acggcattaccaccaaggttaataaaggtgtggtgctgtgtggcattgaccaaaaccaaag 4139  
Db 4081 ACGGCATTACCACCAAGGTAAATAAAGGTGTGTCGTGGCATTTGACCAACCAAG 4140  
QY 4140 gcttaacacgcctaagctgaccgtgggttaataataatggcaaaagcattgtcattgaca 4199  
Db 4141 GCTTAACACACGCTTAAGCTGACCGTGGGTAAATAATATGCGAAAGGCATTTGTCATTGACA 4200  
QY 4200 gccaaaatgggtcaaaaataccatcacagactaaagcaaacacttagctaatgttacaatg 4259  
Db 4201 GCCAAAATGGTCAAAATACCATCACAGACTTAAGCAACACTCTAGCTAATGTTACCAATG 4260  
QY 4260 ataaaggtagcgtacgcacacagacaggggcaataataatcaaaagacgaagaacaaacc 4319  
Db 4261 ATAAAGGTAGGTAGCGACACAGCAAGGCAATATAATCAAGACGAAGACAAACCC 4320  
QY 4320 gtgcgcgcagcattgttgcgttaagcagcggtttaaacttgcgaagcaatggtgaag 4379  
Db 4321 GTGCCGCCAGCATTTGCTGCTTAAGCGCAGGCTTTAACTTTGAAGGCAATGTTGAAG 4380  
QY 4380 cggttgacttctgtccacttatgacacgcgtcaacttggcagatgcaatgccacacg 4439  
Db 4381 CGGTGACTTTGCTTCCACTTATGACACCGTCAACTTTGCCGATGGCAATGCCACACCG 4440  
QY 4440 ctaagtgacctatgatgacaagaacaaacagtaaaagtgctctatgatgtcaatgtgg 4499  
Db 4441 CTAAGGTGACCTATGATGACACAAGCAAAACCCAGTAAAGTGTCTATGATGTCAATGTGG 4500  
QY 4500 atgatacaacacattgaagttaaaagataaaaaacttggcgtaaaaaacacacacattgaca 4559  
Db 4501 ATGATACAAACCATTTAAGTTAAAGATAAAAACTTTGGCGTAAAAAACCCACCATTTGACCA 4560  
QY 4560 gtaactggcacaggtgtctaataatttgcctaaagcaatacaagctactggcgatgcgcttg 4619  
Db 4561 GTACTGGCAGGTGCTAATAAATTTGCCCTAAGCAATCAAGCTACTGCGGATGCGCTTG 4620  
QY 4620 tcaaggccagtgatactgttgcctatcaaacaccttatctgtggcgacatccaaactgcca 4679  
Db 4621 TCAAGGCCAGTGATATCGTTGCTCATCTAAACACCTTATCTGCGGACATCCAACTGCCA 4680  
QY 4680 aaggggcaagccaagcgaacactcagcaggtatgtggatgctgatggcaataagggtca 4739  
Db 4681 AAGGGCAAGGCCAAGCGAACAACCTCAGCAGGCTATGTGGATGCTGATGGCAATAAGGTCA 4740  
QY 4740 tctatgacagtaccgatacaagtaactatcaagccaaaaatgatggcacagttgataaaa 4799  
Db 4741 TCTATGACAGTACCCGATAAACAAGTACTATCAAGCCAAAAATGATGGCACAGTTGATAAAA 4800  
QY 4800 ccaagaagttgccaagaacaaactggtcgcccaagcccaaacccacagatggcacattgg 4859  
Db 4801 CCAAGAAGTTGGCAAAAGACAACCTGGTCGCCCAAGCCCAACCCAGATGSCACATTGG 4860  
QY 4860 ctcaaatgaatgtcaaaaatcagtcattacaagaagaacagtaataatgatgccataaaaagc 4919

Db 4861 CTCAAATGAATGTCAAAATCAGTCATTAACAAAGAACAGTAAATGATGCCAATAAAAGC 4920  
Qy 4920 aagcatcaatgagcaaacgcttctgttaaaggacttgaaagacccgcttctgtataaca 4979  
Db 4921 AAGGCATCAATGAAGCAACAGCGCTTTGTTAAAGGACTTGAAAGAGCGCTTCTGATAACA 4980  
Qy 4980 aaacaaaacgcccagtaactgtgtgtgatttaaatccggttgcacaaacacgcgtga 5039  
Db 4981 AAACCAAAACGCGGAGTAACCTGTGGGTGATTTAAATCCCTTGCCCAACACCGCTGA 5040  
Qy 5040 ccttcaggggatacaggcacacagcgtataaaactggcgagactttgacatacaag 5099  
Db 5041 CCTTTCAGGGGATACAGGCACAAACCGCTTAAACAACTGGCGGAGACTTTGACCATCAAG 5100  
Qy 5100 gtgggcaaacagacacaataaagctaacgcgataataacatcggtgtggtagcaggtactg 5159  
Db 5101 GTGGCAACAGACACCAATAAGCTAACCGATTAATACATCGGTGTGGTAGCAGGTACTG 5160  
Qy 5160 atggttcaactgtcaaacctgtccaaagactaaccaatcttaacagcgttaatgcagggtg 5219  
Db 5161 ATGGCTTCACTGTCAAACTTGGCCAAAGACCTAACCAATCTTAACACGCTTAATGCAGGTG 5220  
Qy 5220 qcacaaaattgatgacaaaggcgtgtctttttagactcaagcgggtcaagccaaagcaa 5279  
Db 5221 GCACCAAAATGTAGCAAAAGCGGTGTCTTTTCTAGACTCAAGCGGTCAAGCCAAAGCAA 5280  
Qy 5280 acacccctgtcttaagtgcacaaatgggctggacctgggtggcaagggtcaatcagtaagtgg 5339  
Db 5281 ACACCCCTGTGCTAAGTGGCAATGGGCTGGACCTGGGTGGCAAGGTCACTCAGTAATGTGG 5340  
Qy 5340 qcaaagcacaaaagataccgacgctgccaatgtacaacagtttaaacgaagtagcgaact 5399  
Db 5341 GCAAGGCACAAAGATACCGACGCTGCCAATGTACAAACAGTTTAACGAGTACGCAACT 5400  
Qy 5400 tgttggtcttggtaagtctgtgttaatgaataacgctgacggcaatcagggtataacatgccc 5459  
Db 5401 TGTGGGTCTTGTGTAATGCTGTAATGATTAACGCTGACGCAATCAGGTAACATGGCC 5460  
Qy 5460 acatcaaaaagaccacaaatccaggttcatcatcaccgcactgtcatcaaacagcagca 5519  
Db 5461 ACATCAAAAGACCCCAATTCAGGTCATCCTTAACCCACTGTCTCAAGACAGGCA 5520  
Qy 5520 cggctacttggcgttaagggtataacgataaccgaaaaacttgcacttgggtatatacag 5579  
Db 5521 CGGTACTTGGCGGTAAAGGTAAATTAACGATACCGAAAAACTTGCACCTGGTGGTATACAAG 5580  
Qy 5580 tggcggtgataaagacggcaacgtaacggaatttaagcaatgtttgggtcaaaaccc 5639  
Db 5581 TGGCGGTGGATAAAGACGGCAACGCTAACGGCGATTTAAGCAATGTTTGGGTCAAAACCC 5640  
Qy 5640 aaaaagatggcagcaaaaagccctgctcgccacttataacgcgcgaggtcagaccaact 5699  
Db 5641 AAAAGATGGCAGCAAAAACCCCTGCTCGCCACTTATAAGCGCGCAGGTACAGACCAACT 5700  
Qy 5700 atttgaccaacaccccgagaaagccattgacagaaataatgaacagggtatccgcttct 5759  
Db 5701 ATTTGACCAACACCCCGCAGAAAGCCATTGACAGATAAATAAGTAAGGATATCCGCTTCT 5760  
Qy 5760 tccatgtcaacgatggcaatcaagcctgtgttacaaggcgtlaacggcattgactcaa 5819  
Db 5761 TCCATGTCAACGATGGCAATCAAGAGCCTGTGTACAAAGGGCGTACGGCATTCACATCAA 5820  
Qy 5820 gtgcctcaggcaagcactcagtggcgatagtttccaggccaaaggcagatgggtgaacccg 5879  
Db 5821 GTGCCCTCAGGCAAGCAGCTCAGTGGGATAGGTTTCCAGGCCAAGCAGATGTTGAAGCCG 5880  
Qy 5880 ccgttgccataggcagacaaacccaaagcaggcaacccaatccatcgccatcggtgataacg 5939  
Db 5881 CCGTTGCCATAGGCAGACAAACCCAAAGCAGCAACCAATTCATCGCCATCGGTGATACG 5940  
Qy 5940 caaagccacagggcgatcaatccatcgccatcggtacaggcaatgtgtagcaggtgaagc 5999  
Db 5941 CACAAGCCACGGCGCATCAATCCATCGCCATCGGTACAGGCAATGTGTTACAGGTAAAGC 6000

RESULT 3

US-08-621-944-1

; Sequence 1, Application US/08621944

Qy 6000 actctggtgccaatcggcgaccccaagcactgttaaggctgataaacagttacagtggtgggta 6059  
Db 6001 ACTTGGTGGCCATCGCGACCCCAAGCACTGTTAAGGCTGATACAGTTTACAGTTTACAGTGGGTA 6060  
Qy 6060 ataaacaacagttaccgatgccactcaaacccgatgtcttttgggtggtggcaataacatca 6119  
Db 6061 ATAACAACCACTTTACCGATGCCACTCAAAACCGATGTCTTTGGTGTGGCAATAAACATCA 6120  
Qy 6120 ccgtgacccgaagaagtaactcgggttgccttaggttcaaaactcgtccatcagtgccagcacac 6179  
Db 6121 CCGTGACCGAAAGTAACCTCGGTTGCCCTTAGGTTCAAACTCTGCCATCTAGTGCAGGCACAC 6180  
Qy 6180 acgcaggcacacaagcccaaaaaactctgcggcacagcaggtlacaaacccacacagcaggtg 6239  
Db 6181 AGCAGGACACACAGCCCAAAAATCTGACGGCACAGCAGGTACAAACACCACAGCAGGTG 6240  
Qy 6240 caaccggctacgggttaaaaggcttctgtgacaaaacggcggttgggtcgtctccgtgggtg 6299  
Db 6241 CAACCGGTACGGTTAAAGGCTTGTGTCGACAAACACGCGGTTGGTGCAGTCTCCGTGGGTG 6300  
Qy 6300 cctcagggtgctaaacgcgcgtatccaaaatgtgcagcaggtgaggtcagtgccaccagca 6359  
Db 6301 CCTCAGGTGCTGAACGCGGTATCCAAAATGTGCGACAGGTGAGGTGAGTGCACCAAGCA 6360  
Qy 6360 ccgatgcggtcaatggttagccagttgtacaaaggccacccaaagcatttgccaaacgcaaca 6419  
Db 6361 CCGATCGCGTCAATGGTAGCCAGTTGTACAAAGCCACCCAAAGCATTGCACACGCAACCA 6420  
Qy 6420 atgacttgacacatcgctatccacaaaacgaaaaataaggccaaatcagggtatctcatcag 6479  
Db 6421 ATGAGCTTGACCATCGTATCCCAAAAGCAAAATAAGGCCAATGCAGGGATTTTCATCAG 6480  
Qy 6480 cgaatgcggtggcgtccatgccaagcctacattctcggcagatccatcaggttacccggg 6539  
Db 6481 CGATGCGGATGGCTCCATGCCACAAAGCCTACATCTCTGCAGATCCATGTTACCGGGG 6540  
Qy 6540 gtattgccaccacacacggtcgaagggtcgggtggcaggtgggactgtcgaagcgtcgggata 6599  
Db 6541 GTATTGCCACCCACACACGCTCAAGGTGCGGTGGCAGTGGGACTGTGCAAGCTGTGCGGATA 6600  
Qy 6600 atggtcaatgggtattataaaatcaatggttcagccgatacccaagccatgtaggggcyg 6659  
Db 6601 ATGGTCAATGGGTATTTAAATCAATGTTTACGCCCATACCCCAAGGCCATGTAGGGGCGG 6660  
Qy 6660 cagttggtgcaggttttccacttttaaggccataaaatcgaagaatttttacttaaaaaatcaat 6719  
Db 6661 CAGTTGTTGCAGGTTTTCATCTTTTAAGCCATAAATCGCAAGATTTTACTTAAAAATCAAT 6720  
Qy 6720 ctacacatagttgtataaaaacagcatcagcatcagcatattactatgctgatgtttt 6779  
Db 6721 CTCACCATAGTTGTATAAAACAGCATCAGCATCAGTCATATTACTGATGCTGATGTTTTT 6780  
Qy 6780 tataccttaaacattttaccgctcaagtgtattctcttccaccatgaccaaatacgcatt 6839  
Db 6781 TATCACTTAACCATTTTACCGCTCAAGTGTATCTCTTTCACCATGACCAAAATCGCCATT 6840  
Qy 6840 gatcataggtaaaccttatgtagtaaattttatcaatgtagttttagatataggttaaaat 6899  
Db 6841 GATCATAGGTAACCTTATTCAGTAAATTTTATCAATGTAGTTTGTAGATATGTTTAAAAAT 6900  
Qy 6900 tgtgcaattgacaaaataagcaggtattatccgaaaaattctctgattgatccgttga 6959  
Db 6901 TGTGCCAATTGACCAAAAATAAGCCGATTATCCGANAATTTCTGATTATGATCCGTTGA 6960  
Qy 6960 cctgcaggtcgac 6972  
Db 6961 CCTGCAGTCCGAC 6973







Db 1381 TTGCCCTAGGTCAGAGTTCTGTGTGTCACCTAGAGTGATATAATTTCTAGACGGCGCTATA 1440  
QY 1440 caccataatcaccaggaacatagaccccaagtttcaagccacaaataataatgaagcggttc 1499  
Db 1441 CACCAAAATCCAGGACATAGACCCCAAGTTTCAAGCCACCAATATAGAAAGCGGGTC 1500  
QY 1500 cactttccatttgtagtaactctatcaaacgtaaaatcatcaatgctgggtaggtgtta 1559  
Db 1501 CACTTTCATTTGGTAGTAATCTATCAACAGTAAATCATCATCTGCGGTGAGGTGTA 1560  
QY 1560 ataaacccgatcggttcaatgtggcacagctagaaagcggtgtgaagtggtgtaaggaac 1619  
Db 1561 ATAAACCGATGCGGTCAATGTGGCACAGCTAGAAAGCGGTGTAAGTGGCTAAGGAGC 1620  
QY 1620 gtgaataattactttcaggtgtgataacagctactgaactgaactaaataagtttggataata 1679  
Db 1621 GTAGAATTTACTTTTCAGGGTGATGATTAACAGTACTGACGTAAATAATAGGTTGGATAATA 1680  
QY 1680 cttaactattaaaggtgtgtgagagacaaacgattaaaccgataataatcgggtgtg 1739  
Db 1681 CTTTAACTATTAAAGGTGGTGACAGACCAACGCATTTAACGGATATAATATCGGTGTGG 1740  
QY 1740 taaagaggtgtataatagttgtctgaagtttaactgtctaaactttaaacaacttta 1799  
Db 1741 TAAAGAGGCTGATATATAGTGGTCTGAAAGTTTAAACTTGTCTAAACCTTTAAACAATCTTA 1800  
QY 1800 ctgaagtgataacactacattaaatgaacacacacagcttaaggttaggttagtagta 1859  
Db 1801 CTGAGGTGAATACAACTPACATTTAAATGCCACACACAGTTAAAGTTAGTTAGT 1860  
QY 1860 gtactacagctgaattattgagtgatgttttaacctttaccagcccaataacagggcagtc 1919  
Db 1861 GTACTACAGCTGAATTTATGAGTGATAGTTTAACTTTTACCAGCCCAATACAGGCAGTC 1920  
QY 1920 aaagcacaagcaaacagctctatggtcgttaaggtgaagtttacttaataatgcagaaa 1979  
Db 1921 AAGCACAAGCAAAACCGTCTATGGCGCTTAAATGGGTGAAAGTTTACTTAATTAATGCGAATA 1980  
QY 1980 caacagcagcaatcggcactactcgtattaccagagataaaattggcttggctcagatg 2039  
Db 1981 CAACAGCAGCAATCGGCACACTCTGATCTGATTTACCAGAGATAAAATTTGGCTTTGCTCGAGATG 2040  
QY 2040 gtgattgttgataaaacaagcaccatatttggataaaacaacttaaggttaggtg 2099  
Db 2041 GTGATGTGTGATGAAAAACAAGCACCATTTTGGATAAAAAACAACCTTAAAGTGGGTAGT 2100  
QY 2100 ttgcaattaccatagacaattggtcattgagtcaggttaataaaagatcagtaaatgttgc 2159  
Db 2101 TTGCNAATACCATAGACANTGGCATTTGATGAGTAAATAAAGATCAGTAATCTTGCCA 2160  
QY 2160 aagtagcagtgctaaagatcggtttaccatcgaaacgctcaaacgccaagcctactt 2219  
Db 2161 AAGGTAGCAGTGTCAACGATCGGTTCACATCGAACACAGCTCAAAAGCCGCCAAGCCTACTT 2220  
QY 2220 taaacgagcgcttgcaatcagtcacactcactgaatcagttgagtcgaagtg 2279  
Db 2221 TAAAGCAGAGCGCTGGCATCAGTGTACACACCTACTGAAATATCAATTTGATGCTAAGAGTG 2280  
QY 2280 gcaattggtaccgcccacacttaacacattggcgtgaaacacccagcagcttaacagtgatg 2339  
Db 2281 GCAATGTTACCGCCCAACTTACAACATTTGGGTGAAACCCACCGAGCTTAACAGTGATG 2340  
QY 2340 gcaactgataataattagtttaaggttagtgtaagcaaatagcttagttaccgctg 2399  
Db 2341 CCACATAGTATATAATTTTATGTTAAGGTAGTGTGATGACCAATAGCTTTACCGCGC 2400  
QY 2400 aacatttggaagctctatcaaatgaatgaatcgaacggctgacagtgctctcaaaagct 2459  
Db 2401 AACATTTGGCAAGCTATCTAATTAAGTCAATCGAAGCGCTGACAGTGTCTCTACAAAGCT 2460  
QY 2460 ttaccgttaaagaagaagcagtagtagcgaacgctctacacggtgcttaagaagcagga 2519  
Db 2461 TTACCGTTTAAAGAGAAGCAGATGATGACGCCAACGCCTATCACCGTGGCTTAAAGATACGA 2520

QY 2520 caaaaaatgcggcgagcagtcagcatcttaaaactcaaaaggtanaaaacgggtctaaacgggtt 2579  
Db 2521 CAAAAAATGCCGGCGCAGTCAGCATCTTAAAACTCAAAAGGTAAAAAGCGGTCTAACCGTTG 2580  
QY 2580 ctacaaaaaagataggtaacggttacccttggccttagcgaagatagcgggtctgaccattg 2639  
Db 2581 CTACCAAAAAGATGGTACGGTTACCTTTGGGCTTTAGCCAAAGATAGCGGTCTGACCATTTG 2640  
QY 2640 gcaaaagcaccctaaacaaacgagcttgctgtttaagataaccaagaacaaatcccaag 2699  
Db 2641 GCANAAGCACCTTAACAAACGATGCTTGAATGTTAAAGATACCAACGAACAAATCCAAAG 2700  
QY 2700 tgggtgctaaatgggaatttaattacttaattgaattggttagtaattcccaaggtactggcattg 2759  
Db 2701 TCGGTGCTTAATGGCATTAATTTACTTAATGTAATGCTAGTAATCCAGGTACTGGCATTTG 2760  
QY 2760 caaataccgcctgcgctaccagagataaaattggccttggcgttctgtgaggtgagttg 2819  
Db 2761 CAATACCGCTCGCATTTACAGAGATTAATTTGGCTTTGCTGGTCTGATGGTGCAGTTG 2820  
QY 2820 atacaaaaaaccttctctctgacagacagctacaagttggcaattggttaagaataacca 2879  
Db 2821 ATACAAACAAACCTTTATCTTGATCAAGCAACAGCTACAAGTTGGCAATGTTAAGATTACCA 2880  
QY 2880 aactggcattaaacgaggtggttaaagccatcaacagggctgtcccaacacgtgcctagca 2939  
Db 2881 ACACCTGGCATTTAACGACAGGTGGTAAAGCCATCACAGGGCTGTCCCAACACTTGCCTAGCA 2940  
QY 2940 ttggcagataaagtagcgcgcaacatagaaactgggcaatatacaatcccaagaacaaagcaaat 2999  
Db 2941 TTGCGCATCAAGTAGCCGCACATAGAACTGGCAATACATCCAAAGCAACAAAT 3000  
QY 3000 ccaacgctgccagcattatgatatataataacacagggcttcaacctaaataaataaaca 3059  
Db 3001 CCAACGCTGCCAGCATTAATGATATATTAATACAGGCTTTAACCTTAAAAATAATAACA 3060  
QY 3060 acccattgaacttgcctcacttatgacattgtgactttgcaatggcaatggcaatgcacca 3119  
Db 3061 ACCCATTTGACTTTGCTCCACTTATGACATTTGTTGACTTTGCCAATGGCAATGCCACCA 3120  
QY 3120 cgcacagtaacccatgataccgcttaacaaacaaacaaagtaagtggtatagtgatgaaatg 3179  
Db 3121 CGCCACAGTAACCCATGATACCGTAAACAAACCAAGTAAAGTGGTATATGATGTAATG 3180  
QY 3180 tggatgatacaaccattctaaacgacgtgacataaaataaaacttggcgtcaaaa 3239  
Db 3181 TGGATGATACCAACCATTCATCTAACAGGCACCTGATGCAATTAANAACCTTGGCGTCAAAA 3240  
QY 3240 ccacccaaactgaacaaaaaagtgctaatggtaatacagcaactaaacttaattgtaact 3299  
Db 3241 CCACCAACTGAACAAACAAAGTGTAAATGGTAAATACAGCACTAACTTTAATGTTAACT 3300  
QY 3300 ctgtagtgaagtagcccttggtaacgcaaaagacatcgccgaaaaatctaaacacccctag 3359  
Db 3301 CTAGTGTAGTAAGATGCGCTTGTAAACGCCAAAGACATCGCCGAAAATCTAANACCCCTAG 3360  
QY 3360 ccaagaaattcacacaccacaaagggcacagcagacacccctcaaaaactttaccgtta 3419  
Db 3361 CCAAGGAATTTACACCCACCAAGGACACAGCACCGCCCTTACAAACCTTTACCGTTA 3420  
QY 3420 aaaaggtagatgaaaaataatgtctgtagcgcgaacgcaatcaacggtgggtgcaaaaaa 3479  
Db 3421 AAAAGGTAGATGAAAAATAATGCTGATGACGCCAACGCCATCACCGTGGGTCAAAAAGA 3480  
QY 3480 acgcaataataaagtaacacacccctcaactcaaggtgaaacgggtcttaataataaaa 3539  
Db 3481 ACGCAAAATATCAAGTCAACCCCTTACACTCAAGGTGAAACGGGTCTTTAATATTAAAA 3540  
QY 3540 ccgcaaaaaatggtagcgggttacctttggcattaaacacacaaagcgttcttaagcgcgga 3599  
Db 3541 CCGACAAAAATGTTACGGTTACCTTTGGCATTAACACCAACAGCGGTCTTTAAAGCCGGCA 3600

QY 3600 aaagcacccctaaacgacgggtggtctgtctctataaaaaacccccactggtgtagcgaacaaatccc 3659  
DB 3601 AAAGCACCCCTAAACGACGGTGGCTGTCTATTAAAAACCCCACCTGGTAGCGAAACAAATCC 3660  
QY 3660 aagtcgggtgctgagtgaggttgccaggttaataataaagtggttgtaggtgctg 3719  
DB 3661 AAGTCGGTGTGATGGCGGTGAAGTTGGCAAGGTTTAATAAATGAGTGTGTAGTGTGCTG 3720  
QY 3720 gcatatggtggcacactgcattaccagagatgaaattggctttactgggactaatggct 3779  
DB 3721 GCATGTGATGGCACAACTCGCATTAACAGAGATGAAATTTGGCTTTACTGGGACTAATGGCT 3780  
QY 3780 cacttgataaagaacaccccccaactaagcaaacgagcgaattaaacgaggtggttaaaaaaga 3839  
DB 3781 CACTTGTATAAAGCAAAACCCCACTTAAGCAAGACGGCATTAACGCGAGGTGGTAAAAAGA 3840  
QY 3840 ttaccaacattcaatcaggtgagatgtgccaaaacagcagcatgagtgctgacagggcgca 3899  
DB 3841 TTACCAACATTCATCAGTCAGTGAATTTGCCAAACAGCCATGATGCTGTGACAGCGGCA 3900  
QY 3900 agatttatgatttaaaacccgaacttgaacacaaaatcagcagctactgcccacaaacagcac 3959  
DB 3901 AGATTTATGATTTAAAAACCGNACTTGAAACAAAAATCAGCAGTACTGCCAAAAACAGCAC 3960  
QY 3960 aaaactcattcacagaattctcagtagcagatgaacaaaggttaataacttttcaggttagta 4019  
DB 3961 AAAACTATTACAGNAATTCCTCAGTAGCAGATGAACAAGGTAAATAACTTTACGGTTAGTA 4020  
QY 4020 acccttactcagttatgacacctcaagacctctgctcatcacctttgcaggtgaaa 4079  
DB 4021 ACCCTTACTCCAGTTATGACACCTCAAGACCTCTGATGTATCACCTTTGCAGGTGAAA 4080  
QY 4080 acggcatttaccaccaaggttaataaagggtggtgctggtggcattgacacaaaccaaag 4139  
DB 4081 ACGGCATTACCACCAAGGTAAATAAAGGTGTGGTGGCGATTGACCAAAACCAAAG 4140  
QY 4140 gcttaacacagcctaagctgacctgggtggttaataataatggcaaggcaattgtcattgaca 4199  
DB 4141 GCTTAACACAGCCCTAAGCTGACCGTGGTGAATAATAATGGCAAGGCATTGTCAATTGACA 4200  
QY 4200 gccaaaattggttcaaaataccatcacagactaagcaacactctagcttaagtctaccaatg 4259  
DB 4201 GCCAAATATGGTCAAAATACCATCACAGACTAAGCAACACTCTAGCTAATGTTACCAATG 4260  
QY 4260 ataaaggtagcgtacgcaccacagaacagggaatataatacaaaagacgaacaaacccc 4319  
DB 4261 ATAAAGGTAGCGTACGCCACCACAGAACAGGGCAATATATAAAGACAGAACAAAAACCC 4320  
QY 4320 gtgcgcgcagcattgttgatgctgaagcagcaggtttaacttgcgaaggcaaatggtgaag 4379  
DB 4321 GTGCCGCCAGCATTTGTTGATGTGCTTAAGCGCAGGCTTTTAACCTTGCNAGGCAATGGTGAAG 4380  
QY 4380 cggttgactttgtctccacttatgacacggtcaacttggcgatggcaatgccacacccg 4439  
DB 4381 CGTTTGACTTTGTCTCCACTTATGACACCGCTCACTTTGCCGATGGCAATGCCACCACCG 4440  
QY 4440 ctaaggtgacctatgatgacacagcaaaacagtaaagtggtctatgaatgcaatgtgg 4499  
DB 4441 CTAAGGTGACCTATGATGACACAAGCAAAACCCAGTAAAGTGGTCTATGATGTCAATGTGG 4500  
QY 4500 atgatacaaccaattgaagttaaagataaaaaacttggcgtaaaaaacacacacattgacaa 4559  
DB 4501 ATGATACAAACCATTAAGCTTAAGATNAANAACCTTTGGCGTAAAAACCAACCATTTGACCA 4560  
QY 4560 gtactggcacaggtgcttaataattgccttaagcaatcaagctactggcgatgcgcttg 4619  
DB 4561 GTACTGGCACAGGTGCTAATAATTTGCCCTTAAGCAATCAAGCTACTGGCGATGGCGCTTG 4620  
QY 4620 tcaaggccagtgatcgttgctcatctaaacaccttatcttggcgacatcccaactgcca 4679  
DB 4621 TCAAGGCCAGTGATATGCTGCTCATCTAAACACCTTATCTGGCGACATCAACACTGCCA 4680  
QY 4680 aaggggcgaagccaaagcgaacaaactcagcaggtcctagtgtgctgagtgagtgagtgagtgca 4739

DB 4681 AAGGGCAAGCCAAAGCGAACCAACTCAGCAGGCTATGTGGATGCTGATGCCAATAAGGTCA 4740  
QY 4740 tctatgacagtlaccgataacaagtlactatcaagcccaaaaatgagtcagcttgataaaa 4799  
DB 4741 TCTATGACAGTACCAGTAACAAGTACTATCAAGCCCAAAAATGATGGCACAGTTGTATAAA 4800  
QY 4800 ccaagaagttgtccaaagacaaactggtgcccaagcccaacccccagatggcacattgg 4859  
DB 4801 CCAAGAAGTTGCCAAAGACAACTGGTCGCCCAAGCCCAAAACCCAGATGGCACATTGG 4860  
QY 4860 ctcaaaatgaatgtcaaatcagtcattaaacaagaacaagtaaaatgatgccataaaaaagc 4919  
DB 4861 CTCAATGAATGTCAATCAGTCATTAAACAAGAAACAAGTAAATGATGCCAATAAANAAGC 4920  
QY 4920 aaggcatcaatgaagacaacaccttgttaaaggacttgaaaaagccctctctgataaca 4979  
DB 4921 AAGGCATCAATGAAGACAAACCGCTTTGTTAAAGACTTGAANAAGCCCTCTCTGATAACA 4980  
QY 4980 aacccaaaacgcgcgcagtaactgtgggtgatttaaatgcccgttgcacacacgcgtga 5039  
DB 4981 AAACCAAAAACGCCGACGTAACCTGTGGGTGATTTAAATGCCGTTGCCCAACACCCGCTGA 5040  
QY 5040 cctttgcagggtgatacaggcaacacggtctaaaaaactggcgagactttgacacataaaag 5099  
DB 5041 CCTTTGCAGGGGATACAGGCACACCGGTAAAAAATCGGCGAGACTTTGACCATCAAAG 5100  
QY 5100 gtgggcaacagacacaccaataagctaacccgataataacatcgggtgtgtgacagctactg 5159  
DB 5101 GTGGCAAAACAGACACCAANTAAGCTAAACCGATATAATACCTCGTGTGGTAGCAGGTACTG 5160  
QY 5160 atggtctcactgtcaaaacttgcacaaagaccttaaccaactcttaacagcgttaaatgcaggtg 5219  
DB 5161 ATGGCTTCACTGTCAAACTTGCCAAAGACCTAAACCAATCTTAACAGCGTTAATGCGAGGTG 5220  
QY 5220 gcaccaaataatgatgacaaaagcgtgtctttgttagactcaagcgggtcaagccaaagcaa 5279  
DB 5221 GCACCAAAATATGATGACAAAGCGTGTCTTTGTAGACTCAAGCGGTCAAGCCAAAGCAA 5280  
QY 5280 acacccctgtctgaagtccaatggcttggactggcctgggtggcaagctcatcagtaagtgg 5339  
DB 5281 ACACCCCTGTCTAGTCCCAATGGCTGGACCTGGGTGGCAAGGTCTATCAGTAATGTGG 5340  
QY 5340 gcaaggcacaaaagataccgcagctgccaatgtacaacagttacaacgaagtacgcaact 5399  
DB 5341 GCAAAAGGCACAAAAGATACCGACGCTGCCAATGTACAAGTAAACAGTAACGAAGTACGCAACT 5400  
QY 5400 tgttgggtcttggtaagtgtgtaagctgataacgctgacgggcaatcaggtaaaacattgccc 5459  
DB 5401 TGTGGGTCTTTGGTAATGCTGGTAATGATAGGCTGACGGCAATCAGGTAAACATTTGCCG 5460  
QY 5460 acatcaaaaaagacccaaattcaggttcatcatctaacgcgactgtcatcaaacgacgca 5519  
DB 5461 ACATCAAAAAGACCCAAATTCAGGTTTCATCATCTTAACCGCACTGTCTATCAAAAGCAGGCA 5520  
QY 5520 cggctacttggcggtaaaggtaaataacgataccgaaaaaacttggcacttgggtgtatcaag 5579  
DB 5521 CGGTACTTTGGCGGTAAAGGTAAATACGATACCGAAAAAATTTGCCACTTGTGTGTATACAAG 5580  
QY 5580 tggcggttgataaaagacgcaacgctaacggcgatttaagcaatttgggttcaaaacccc 5639  
DB 5581 TGGCGGTGGATAAAGACGGCAACGCTAACGGCGATTTAAGCAATGTTTGGGTCAAAACCC 5640  
QY 5640 aaaaagatggcagcaaaaaagccctgctgcacactataaacgcccgaggtcagacaaact 5699  
DB 5641 AAAAAGATGGCAGCAAAAAAAGCCCTGCTGCCACTTATAACGCCCGAGGTTCAGACCAACT 5700  
QY 5700 attgacacaaaccccccgagaagccattgacagataaaatgaacaaaggtatccgcttct 5759  
DB 5701 ATTTGACCAACAAACCCCGCAGAAAGCCATTGACAGAATAAATGAACAAGTATCCGCTTCT 5760  
QY 5760 tccatgtcaacgattggcaatcaagagcctgtgtgtacaagggcgtaacggcatttgactcaa 5819

Db 5761 TCATGTCACCATGCAATCAAGAGCCTGTGGTACAGGGCGGTACGGCATGACTCAA 5820  
Qy 5820 GTGCTCAGGCAAGCACTAGTGGGATAGTTCAGGCGCAAGCAGATGTTGAAGCG 5879  
Db 5821 GTGCTCAGGCAAGCACTAGTGGGATAGTTCAGGCGCAAGCAGATGTTGAAGCG 5880  
Qy 5880 CCGTTCATAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 5939  
Db 5881 CCGTTCATAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 5940  
Qy 5940 CCAAGCCAGGCGATCAATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCG 5999  
Db 5941 CCAAGCCAGGCGATCAATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCG 6000  
Qy 6000 ACTCTGTGCTCGCATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCGCAT 6059  
Db 6001 ACTCTGTGCTCGCATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCGCAT 6060  
Qy 6060 ATAACAACCAAGTTACCGATGCCACCAACCAAGCAAGCAAGCAAGCAAGCAAGCA 6119  
Db 6061 ATAACAACCAAGTTACCGATGCCACCAACCAAGCAAGCAAGCAAGCAAGCAAGCA 6120  
Qy 6120 CCGTACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6179  
Db 6121 CCGTACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6180  
Qy 6180 ACAGGCGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 6239  
Db 6181 ACAGGCGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 6240  
Qy 6240 CAACCGGTACGCTAAAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6299  
Db 6241 CAACCGGTACGCTAAAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6300  
Qy 6300 CCGTACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6359  
Db 6301 CCGTACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6360  
Qy 6360 CCGTACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6419  
Db 6361 CCGTACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6420  
Qy 6420 ATGAGCTGACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6479  
Db 6421 ATGAGCTGACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6480  
Qy 6480 CCGTACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6539  
Db 6481 CCGTACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6540  
Qy 6540 GTATGCGCACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6599  
Db 6541 GTATGCGCACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6600  
Qy 6600 ATGAGCTGACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6659  
Db 6601 ATGAGCTGACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6660  
Qy 6660 CCGTACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6719  
Db 6661 CCGTACCAAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6720  
Qy 6720 CTCACCATAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6779  
Db 6721 CTCACCATAGTAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6780  
Qy 6780 TATCACTTAAACCATTTTACCCTCAAGTGAATCTCTTTTACCATGACCAAAATCGCCATT 6839  
Db 6781 TATCACTTAAACCATTTTACCCTCAAGTGAATCTCTTTTACCATGACCAAAATCGCCATT 6840  
Qy 6840 GATCATAGGTAACTTATGAGTAAATTTTATCAATGATGATGATGATGATGATGATGAT 6899  
Db 6841 GATCATAGGTAACTTATGAGTAAATTTTATCAATGATGATGATGATGATGATGATGAT 6900

Qy 6900 TGTGCCATTGACCAAAAATGACCGATTATCCGAAATTTCTGATTATGATCCGTTGA 6959  
Db 6901 TGTGCCATTGACCAAAAATGACCGATTATCCGAAATTTCTGATTATGATCCGTTGA 6960  
Qy 6960 CCTGCAGGTGAC 6972  
Db 6961 CCTGCAGGTGAC 6973

RESULT 4  
US-08-621-944A-1  
; Sequence 1, Application US/08621944A  
; GENERAL INFORMATION:  
; APPLICANT: SASAKI, Ken  
; APPLICANT: HARKNESS, Robin E.  
; APPLICANT: LOOSMORE, Sheena M.  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER  
; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; COUNTRY: Canada  
; STATE: Ontario  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/621,944A  
; FILING DATE: 26-MAR-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/478,370  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-587  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6973 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 708..6683  
; US-08-621-944A-1

Query Match 99.8%; Score 6961; DB 10; Length 6973;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6972; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 CCATGGATATGGGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60  
Db 1 CCATGGATATGGGCGAGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60  
Qy 61 CATATCTGTACGATTTGACATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120  
Db 61 CATATCTGTACGATTTGACATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120



Db 2281 GCAATGTTACCGCCCAACTTACAACATTGGCGTGAATAACCAACCGAGCTTAACACTGATG 2340  
Qy 2340 gcaactagatataaatttagtgttaagggttagtgggtacgaacaataagcttagttaccgcg 2399  
Db 2341 GCACCTAGTATAAATTTAGTGTAAAGGTAGTGGTACGAACAATAGCTTAGTTACCGCGG 2400  
Qy 2400 acatttggcaagctactcaaatgaagtcaatcgcaacggctgacagtgctctacaaagt 2459  
Db 2401 AACATTTGGCAAGCTATCTAATAGAGTCAATCGAAGGGCTGACAGTGTCTTACAAAGCT 2460  
Qy 2460 ttaccgttaagaagaagcagatgataagcgcacaaagctatcacctggtgctaaagatacga 2519  
Db 2461 TTACCTGTTAAAGAAGAAGACGATGATGACGCCAAACGCTATCACCTGGCTGAAGATACGA 2520  
Qy 2520 caaaaaatccgcgcagctcagcatcttaaaactcaaaaggttaaaacgctctaaacggttg 2579  
Db 2521 CAAAAAATGCGCGGCGAGTCAGCATCTTAACACTCAAAAGTAAACGGTCTAACGGTTG 2580  
Qy 2580 ctaccaaaaaagatggtacggtttacaccttgggttagcgaagatagcggctctgaccattg 2639  
Db 2581 CTACCAAAAAAGATGTTACGGTTACCTTTGGGCTTAGCCAAGATAGCGGCTCTGACCATTG 2640  
Qy 2640 gcaaaagcacccttaaacacagatggtctgactgttaagatatccaacgaacaaatccaaag 2699  
Db 2641 GCAAAAGCACCTTAAACACGATGGCTTGACTGTTAAAGATACCAACGAACAAATCCAAG 2700  
Qy 2700 tcggtgctaatggcattaaatttactaaatgaatgtagtaattccaggtactggcattg 2759  
Db 2701 TCGGTGCTAATGCGATTAATTTACTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 2760  
Qy 2760 caaatccgctcgcattaccagagataaaattggtcttgcctggtctgtggtggtggtggtggtg 2819  
Db 2761 CAAATACCGCTCGCATTTACCAGAGATAAATTTGGCTTTGCTGTTGCTGATGGTGCAGTTG 2820  
Qy 2820 atacaaacaaaccttactgtatcaagacaagctacaagttggcaatgttaagattacca 2879  
Db 2821 ATACAAACAAACCTTATCTTGATCAAGACAAGCTACAAAGTTGGCAATGTTAAGATTACCA 2880  
Qy 2880 acactggcattaaacgcagggtggtaagccatcacagggtgtgtcccaaacactgccttagca 2939  
Db 2881 ACCTGCGCATTTAAGCAGGTGGTAAAGCCATCACAGGGCTGTCCCAACACTGCCTAGCA 2940  
Qy 2940 ttgcgctaaagttagcgcgaacatagaaactgggcaatacaatccaagacaagaacaaat 2999  
Db 2941 TTGCGCATCAAGTAGCGGCAACATAGAACTGGGCAATACAAATCCCAAGACAAGACAAAT 3000  
Qy 3000 ccaacgctgcagcatttaattgatatataataacaggttttaaccttaaaataaataaaca 3059  
Db 3001 CCAACGCTGCCAGCATTAATGATATATTAATACAGCTTTAACCTTAAAAAATAATAACA 3060  
Qy 3060 accccattgacttgtctccacttatgacatgttgacttggcaatggcaatggcaatggcaat 3119  
Db 3061 ACCCATTTGACTTTGTCTCCACTTATGACATTTGTTGACTTTGGCAATGGCAATGGCCACCA 3120  
Qy 3120 ccgcacagtaaccatgataccgtcaacaaacagtaagtggtatatgattgtaagt 3179  
Db 3121 CGGCCACMGTAACCCATGATACCGCTAACAAACCACTAAGTGGTATATGATGTAATG 3180  
Qy 3180 tggatgatataccacttcaatcaacagcactgatgacaaataaaaaacttggcgctcaaaa 3239  
Db 3181 TGGATGATACAACCATTCATTAACAGGCACCTGATGACATTAATAAATAAATAAATAAATAA 3240  
Qy 3240 ccacaaactgaacaaacaaagtctaatgtaatacagcaactaatcttaattgtaact 3299  
Db 3241 CCACCAAACTGAACAAACAAGTGTAAATGTAATACAGCACTAACCACTTAATGTTAACT 3300  
Qy 3300 ctatgataagatgaccttggtaacgcgaagacatcgccgaataatcctaaccacctag 3359  
Db 3301 CTAGTGTAGAGATGCCCTTGTAAAGGCCAAAGACATCGCGGAAATCTTAAACACCTAG 3360  
Qy 3360 ccaaggaaattcacaccccaaaaggcacagacacccgacctcaaaacctttaccgtta 3419  
Db 3361 CCAAGGAAATTCACACCACCAAGGACAGACACCGCGCTACAAACCTTTACCGTTA 3420

Qy 3420 aaaaggtagatgaataataatgtctgatgacgcacaacgcatcacctgggtggtcaaaaga 3479  
Db 3421 AAAAGGTAGATGAATAATTAATGCTGATGATGCGCAACGCCATCACCGTGGGTCAAAAAGA 3480  
Qy 3480 acgcaataatacaagtcacacccctaacactcaaaagtgaaaaacggtcttaattattaaaa 3539  
Db 3481 ACGCAAAATTAATCAAGTCAACACCTTAACACTCAAAAGGTGAACCGGTCTTAATATTAAAA 3540  
Qy 3540 ccgcaaaaaatggttacggtttacctttggcattaacaccacaagcggctctaaagccggca 3599  
Db 3541 CCGACAAAAATGCTACGGTTACCTTTGGCATTAACACCACAACGGTCTTAAGCCGGCA 3600  
Qy 3600 aaagcaccctaaagcagcgggtggtctgtctataaaaaacccacacbtggtgagcaacaaatcc 3659  
Db 3601 AAAGCACCTTAACAGCAGGTGGCTTGTCTATTAAAAACCCCACTGCTAGCGAACAATATCC 3660  
Qy 3660 aagtcgggtgctgtagcgtggaagtttccaaggttcaataataatggttcttagtgctg 3719  
Db 3661 AAGTCGGTGTCTGATGGCGTGAAGTTTGGCAAGGTTAAATAAATGCTGTTGTAGTGCTG 3720  
Qy 3720 gcatlgaaggcacaactcgcattaccagagatgaattggctttactgggactaatggt 3779  
Db 3721 GCATTTGATGGCACAACTCGCATTAACAGAGATGAATAATGGCTTTACTGGGACTAATGGCT 3780  
Qy 3780 cacttgaataaagaacaaacccacctaagcaagaacggtcatttaacgcaggtggttaaaaaaga 3839  
Db 3781 CACTTGTNTAAAAAGCAACCCCACTTAAGCAAAAGACGGCATTAACGCAGGTGGTAAAAAGA 3840  
Qy 3840 ttaccaacattcaatcaggtgagattgcccacaaacagcagcagcagcagcagcagcagcagc 3899  
Db 3841 TTACCAACATTCATCACTAGGTGAGATTGCCCAAAACAGCCATGATGCTGTGACAGCGGCA 3900  
Qy 3900 agatttatgatttaaaacccgaacttgaacaaacaaatcagcagtagctgccaaaaacagcagc 3959  
Db 3901 AGATTTATGATTTAAAAACCGAACTTGAACAAACAAATTCAGCAGTACTGCCAAAACAGCAC 3960  
Qy 3960 aaactcattacacaattctcagtagcagatgaagaagtaataacttaccggttagca 4019  
Db 3961 AAAACTCATTAACAGCAATTTCTCAGTAGCAGATGAACAGGTAACTTAACCTTACGTTAGTA 4020  
Qy 4020 accttactccaggttatgacacactcaaacactcagcagcagcagcagcagcagcagcagcagc 4079  
Db 4021 ACCCTTACTCCAGTTATGACACTCAAGACCTCTGATGTCATCACCTTTGCGAGGTGAAA 4080  
Qy 4080 acggtattaccacccaaggttaataaaaggtggtgctgctggcagcagcagcagcagcagcagc 4139  
Db 4081 ACGGCATTTACCAACCAAGGTAATAAAGGTGGTGGGTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 4140  
Qy 4140 gcttaaccgcgcctaaagctgacccgtgggtaataataatggcaaaagcagcagcagcagcagc 4199  
Db 4141 GCTTAACCAAGCCTTAAGCTGACCGTGGTAAATAATAATTAAGCAAAAGCAATGTCATTGACA 4200  
Qy 4200 gccaaaaatggtcaaaaatacactcaacagcagcagcagcagcagcagcagcagcagcagcagc 4259  
Db 4201 GCCAAAAATGGTCAAAAATACCATCACAGGACTAAGCAACACTCTAGCTAATGTTACCAATG 4260  
Qy 4260 ataaaggtagcgtacgcacacacagacagggcaataataataaaagcagcagcagcagcagcagc 4319  
Db 4261 ATAAAGGTAGCGTACGCCACACACAGCGCAATATATCAAAAGCAGCAAGACAAAACCC 4320  
Qy 4320 gtgcccacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4379  
Db 4321 GTGCCGCCAGCATTTGTGATGTGCTAAGCGCAGGCTTTAACTTGCAGGCAATGSGTGNAG 4380  
Qy 4380 cgggtgacttggctccacttatgacacgtcaacttccgagtgcaatgcccacccagc 4439  
Db 4381 CGGTTGACTTTGTCTCCATTATGACACCGCTCAACTTTCGCCGATGGCAATGCCACCCAG 4440  
Qy 4440 ctaaggtgacattatgacacacagcaaaacccagtaaaagtggtcttatgattgcaatggtg 4499  
Db 4441 CTAAGGTGACCTATGATGACACAAAGCAAAACAGTAAAGTGGTCTATGATGTCATGTCATGTCG 4500

QY 4500 atgatacaaacattgaagttaaaataaaaaaacttgccgttaaaaaaacccaccattgacca 4559  
Db 4501 ATGATACAAACATTGAAGTTAAAGATAAAAAAATGGCGTTAAAAACCCACCATGACCA 4560  
QY 4560 gtactggcacaggctgcttaataattgcccctaaagcaatcaagctactgagcgatcgcttg 4619  
Db 4561 GTACTGGCACAGGTGCTTAATAAATTTGCCCTTAAGCAATCAAGCTACTTGGCCGATGCGCTTG 4620  
QY 4620 tcaaggccagtgatctgttctcatctataacaccttactctgggacatccaaactgcca 4679  
Db 4621 TCAAGGCCAGTGATATCGTTCTCATCTAAACACCTTATCTGGCGACATCCAAACTGGCA 4680  
QY 4680 aaggggcaagccaaagcaaacactcagcaggctatgtgtagctgtagtggaataaagtgca 4739  
Db 4681 AAGGGCAAGCCAAAGCGAACAACTCACAGGCTATGTGGATGTCTGATGGCAATAAGGTCA 4740  
QY 4740 tctatgacagtaccgataaacaagtactatcaagccaaataatgatggcacagttgatataaa 4799  
Db 4741 TCTATGACAGTACCAGATANCAAGTACTATCAAGCCAAAATGNTGGCACAGTTGATAAAA 4800  
QY 4800 ccaagaagttgcccagaacaaactggtgcccaagcccaaaccccgatggcacattgg 4859  
Db 4801 CCAAGAAAGTTGCCAAAGACAACTGGTGCGCCAAGCCCAACCCAGATGSCACATTGG 4860  
QY 4860 ctcaaatgaatgtcaaatcagctatcaatacaagaagaagtaataatgatcccaataaaaagc 4919  
Db 4861 CTCAAATGAATGTCAAAATCAGTCAATTAACAAAGAACAAAGTAAATGATGCCAATAAAAAAGC 4920  
QY 4920 aaggcatcaatgaagacaacgcctttgttaaggacattgaagaagccgcttctgataaca 4979  
Db 4921 AAGGCATCAATGAAGACAAACGCCCTTTGTTAAGGACTTGAAAAAGCCGCTTCTGATAACA 4980  
QY 4980 aacccaaaacgcgcagtaactggtgatttaaatgcccgttgcccaaacacccgctga 5039  
Db 4981 AAACCAAAAACGCCGAGTAACCTGTGGGTGATTTAAATGCCGTTGCCCAACACCGCTGA 5040  
QY 5040 cctttgcaagggatataaggaacacagcgtcaaaaaactggggagactttgaccatcaag 5099  
Db 5041 CCTTTGCAAGGGGATACAGGCAACACGGCTAAAAAATCGGGGAGACTTTGACCATCAAG 5100  
QY 5100 gtgggcaaacagacacaataatgaacttaacccgataataacatcggtgtgtagcaggtactg 5159  
Db 5101 GTGGCAAAACAGACACCAANTAAGCTAACCGATATATACTCGGTGTGTGATCAGGTACTG 5160  
QY 5160 atggttctactgtcaaaacttgcgaagaccctaaaccaactcttaacagcaggttaatgcaggtg 5219  
Db 5161 ATGGCTTCACTGTCAAACTTGCCAAAGACCTTAACCAATCTTAACAGCGTTTAATGCAGGTG 5220  
QY 5220 gcaccaaataatgatgacaaaagcgctgtctttgttagactcaagcgggtcaagccaaaagcaa 5279  
Db 5221 GCACCAAAATTTGATGACAAAGGCGTGTCTTTTGTAGACTCAAGCGGTCAAGCCAAAAGCAA 5280  
QY 5280 acacccctgtctgaagtgcgaatgggctggacctgggtggcgcaaggtcatcagtaagtgg 5339  
Db 5281 ACACCCCTGTCTAAGTGCCTAANTGGCTTGACCTGGGTGGCAAGGTATCATAGTATATGTGG 5340  
QY 5340 gcaaggccacaaaagataccgacgtgcgaatgtacacagtttaaacgaagtacgcaact 5399  
Db 5341 GCAAAAGGCACAAAAGATACCGACGCTGCCAATGTACAACAGCTTAACAGCACTACGCAACT 5400  
QY 5400 tgttgggtcttggtaagtctgtaataatgataacgctgaagcgaataatcaggttaacattgccc 5459  
Db 5401 TGTTTGGGTCTTTGGTAATGCTGGTAATGATAACGCTGACGGCAATCAGGTAACAACTTGGCG 5460  
QY 5460 acatcaaaaaagacccaaattcaggttcatcatctaaccgcactgtcatcaaaagcagcca 5519  
Db 5461 ACATCAAAAAGACCCCAANTTCAGGTTTCATCTTAACCGCACCTGTCTCAAAAGCAGGCA 5520  
QY 5520 cggctacttggcgttaaggttaataacgataccgaaaaaacttgccactgttggtatatacaag 5579  
Db 5521 CGGTACTTTGGCGGTAAGGTAAATACGATACCGAAAAAATTTGCCACTTGGTGTATACAG 5580  
QY 5580 tgggcgtggataaagacgcgaacgtaacgcgagtttaagcgaatgtttgggttcaaaaacccc 5639

Db 5581 TGGCGTGGATAAAGACGGCAACGCTAACGGCGATTTAAGCAATGTTTGGGTCAAAACCC 5640  
QY 5640 aaaaagatggcagcaaaaaagccctctgctccactataacgcgcaggtcagaccaact 5699  
Db 5641 AANAAGATGGCAGCNCNAAAAGCCCTGCTCGCCACTTATAACGCCGAGCTCAGACCAACT 5700  
QY 5700 attgacaaacaaaccccgacagaagccattgacagaataaataagaacaaaggtatccgcttct 5759  
Db 5701 ATTTGACCAACAACCCCGCAGAAAGCAATTGACAGAATAAATGAACAAGGTATCCGCTTCT 5760  
QY 5760 tccatgtcaacgatggcgaatcaagagccgtgtgtgtacaagggcgtaaacggcattgactcaa 5819  
Db 5761 TCCATGTCAACGATGGCAATCAAGAGCCTGTGGTACNAGGGCGTAACGGCATTTGACTCAA 5820  
QY 5820 gtgcctcaggcaagcaactcagtgccgataggtttccaggccaaagcagatggtgaagccg 5879  
Db 5821 GTGCTCAGGCACAGCACTCAGTGGCGATAGGTTTCCAGGCCNAGGCAGATGTTGAAGCG 5880  
QY 5880 ccgttgccataggcagacaaaaccccaagcaggaacccaatccatcgccatcggtgataacg 5939  
Db 5881 CCGTTGCCATAGGCAGACAAAACCCCAAGCAGCAACCAATCCATCCCATCGGTGATAAG 5940  
QY 5940 cacaagccacgggcgatcaatccatcgccatcggttacaggccaatgtgttagcaggtaagc 5999  
Db 5941 CACAAGCCACGGCGGATCAATCCATCGCATCGGTACAGGCAATGTGTTAGCAGGTAAGC 6000  
QY 6000 actcgttgccatcggcagcccaacagcactgtaagctgataaagctacagttacagttgggta 6059  
Db 6001 ACTCTGGTGCCATCGGCCACCCCAAGCACTGTTAAGGCTGTAAACAGTTTACAGTGTGGGTA 6060  
QY 6060 ataacaacagctttaccgatgcccactcaaacccgagtctcttgggtgggcaataaacaaca 6119  
Db 6061 ATAACAACCACTTACCGATGCCACTCAACCGGATGTCTTTGGTGGCAATAACATCA 6120  
QY 6120 ccgtgacgcgaagtaactcgtgtgcttaggttcaaacctctgccaatcagtcagtcagcacac 6179  
Db 6121 CCGTGACCGAAGTAACCTCGGTGCTTGGTTTAAAGCTTTCAGGCTGCTCCATCATGTCAGCACAC 6180  
QY 6180 acgcaggcacacaagccaaaatactgacgcacagcaggttacaacccaccacacagcagtg 6239  
Db 6181 ACGCAGGCACACAGCCAAAATACTGACGGCACAGCAGGTACAACACCACAGCAGCGGTG 6240  
QY 6240 caacccgttacggttaagcgttctggtgacaaaacggcgttgggtgcttccctggtggtg 6299  
Db 6241 CAACCGGTACGGTTAAAGGCTTTGCTGGCAAAACGGCGTGGTGGTGGGTGCTCCGTGGGTG 6300  
QY 6300 cctcaggtgctgaacgcggtatccaaaatgtggcagcaggtgaggtcagtcagtcaccacga 6359  
Db 6301 CCTCAGGTGCTGAACGCCGTATCCAAAATGTGGCAGCAGGTGAGTGCACACAGCA 6360  
QY 6360 ccgatgcggtcaatggttagccagttgtacaagcccaacaaagcattgccaacgcgaaca 6419  
Db 6361 CCGATGCGGTCAATGGTAGCCAGTGTACAAGCCACCCCAAAAGCAATGCCAACGCCAACCA 6420  
QY 6420 atgagcttgacatcgatccacacaaaacgaaaaataaaggccaatgcagggatttccatcag 6479  
Db 6421 ATGAGCTTGACCATCGTATCCACCAAAAACGAAAAATAAGGCCAATGCAAGGATTTCATCAG 6480  
QY 6480 cgaatggcgatgggtccatgcccagaagcctacattccttggcagatccatggttaccgggg 6539  
Db 6481 CGATGGCGATGGCGTCCATGCCAACAGCTACATTTCTTGGCAGATCCCATGTTTACCGGG 6540  
QY 6540 gtaattgccaccccaaacgcgtcaagdtcggtggcagtggtgagctgcgaagctgctcggtata 6599  
Db 6541 GTATTGCCACCCCAACCGGTCAAGGTGGGTGGAGTGGGACTGTGGAAGCTGTGCGGATA 6600  
QY 6600 atggtcgaatgggtattttaaatcaatggttccagccgatlaccgaagccatgtagggggcg 6659  
Db 6601 ATGGTCAATGGGTATTTAAATCAATGTTTACGCCGATACCCAAAGGCCATGTAGGGGCGG 6660  
QY 6660 cagttgggtgcaggttttccacttttaagccataaatacgaagaatttcttaataaataatcaat 6719







Qy	1080	ttaacggccatgcagttattaaagaataacgaagctcaaaggataatgatgtataatata	1139
Db	1081	TTAACGCCATGCAGTATTAAAAGAAATACGAAGCTCAAGAGTAATGATGTAAATATA	1140
Qy	1140	gacgcacaaccgaagcgagacgcagctactgcagtggaagccatctcatatgcacagg	1199
Db	1141	GACGCACAACCGCAAGCGACACGCCAGTACTGCAGTGGGAGCCATGCTATATGCACAGG	1200
Qy	1200	gtcatitttccaacgcctttggtacacgggcaacagctaaagtgccctattctcttgccag	1259
Db	1201	GTCAATTTTCCAAGCCTTTTGGTACACGGCAACAGCTAAAGGTGCCTATTCTCTGGCAG	1260
Qy	1260	tggtgttctgccgccacagccggaggccaaatctacaatogctattggtctctgatgcacaat	1319
Db	1261	TGGGTCTTGCCGCACAGCCGAGGCCAATCTACAATCGCTATTGGTCTCGATGCACCAT	1320
Qy	1320	ctagctcggttggaagcgtatagcccttgctgcaggtactcgtgctccagctcacaggcagta	1379
Db	1321	CTAGCTGCTGGGAGCGATAGCCCTTGGTGCAGGTACTCGTCTCAGCTACAGGGCAGTA	1380
Qy	1380	ttgccctagtgcaaggttctgttgcactcagagtgaataaattctagaccggccctata	1439
Db	1381	TTGCCCTAGSTCAAGGTTCTGTGTCACTCAGAGTGATAATAATTCTAGACCGSCTATA	1440
Qy	1440	caccaaataccaggcaactagaccacccagtttcaagccaccacaaataacgaagcgggtc	1499
Db	1441	CACCAAATACCCAGGCACCTAGACCCCAAGTTTCAAGCCACCAATAATACGAAGCGCGGTC	1500
Qy	1500	cactttccattggtagtaactctatacaacgttaaaatcataatgctcggtgcagggttta	1559
Db	1501	CACTTTCATTTGGTAGTAATCTCTATCAACCGTAAAAATCATCAATGTCGGTGCAGGTGTTA	1560
Qy	1560	ataaaaccgatcggtcgaatgtggcacagctagaagcgggtggaagtgggctaaaggagc	1619
Db	1561	ATAAAACCGATCGGTCAATGTGGCACAGCTAGAAGCGGTGCTCAAGTGGGCTAAGGAGC	1620
Qy	1620	gtagaattactttcaagggtgatataacagttactgacgttaaaaatagggtttggataata	1679
Db	1621	GTAGAATTACTTTTCAGGTGTATGATACAGTACTGCGTAAAAAATAGGTTTGATTAATA	1680
Qy	1680	ctttaactataaaagtggtgcagagacaaacgcattaaaccgataaataatcgggtgtgg	1739
Db	1681	CTTTTAACATATAAAGGTGTCGAGACCAACGCATTAAACCGATAATAATATCGGTGTGG	1740
Qy	1740	taaaagagcctgataatagtggtctgaaagttaaacttgtcaaacctttaaaccaacttta	1799
Db	1741	TAAAGAGGCTGATAATAGTGGTCTGAAAGTTAAACTTTGCTAAAACTTTAAACAAATCTTA	1800
Qy	1800	ctgagtgtaatacaactacattaaatgcccacaaccacagtttaaggttaggttagtagta	1859
Db	1801	CTGAGGTGAATACAACTACATTAATATGCCAACCCACACAGTTTAAGGTAGGTAGTAGTA	1860
Qy	1860	gtactacgctgaattcattgagttagtttaaacctttaccagcccaatacacaggcagtc	1919
Db	1861	GTACTACAGCTGAATTTATTTAGTGATAGTTTAACTTTACCCGCCCAATACAGGCAGTC	1920
Qy	1920	aaagcacaagcaaacccgtctatggcggttaatgggggtgaagtttactaataatgcagaaa	1979
Db	1921	AAAGCACAAGCAAAACCGTCTATGGCGTTAATGGGGTGAAGTTTACTTAATAATGCAGAAA	1980
Qy	1980	caacacagcaactcggcactactctattaccagagataaaattggccttgcctcgagatg	2039
Db	1981	CACAGCAGCAATFCGGCATCTCTGTTATACCAGAGATTAATAATTTGGCTTTGCTCGAGATG	2040
Qy	2040	gtgatgttgatgaaaaaacgaccataatttggaataaaaacaaacttaaaagttgggtagtg	2099
Db	2041	GTGATGTTGATGAAAAACAAGCACCATATTTGGATAAAAACAACACTTTAAAGTGGGTAGTC	2100
Qy	2100	ttgcaattacca tagacaatggcatgtatgcaggtaataaaaaagatcagtaattcttgcca	2159
Db	2101	TTGCAATTTACCATAGCAATGGCATTTGATGCAGGTAAATAAAAGATCAGTAATCTTTGCCA	2160

QY	2160	aagtagcagtgctaaacyatgcggttaccatogaacagctcaaaagccgcgaagctactt	2219
DB	2161	AAGGTAGCAGTGCCTAACGATGCGGTTACCATCGAACAGCTCAAAGCGCAAGCCTACTT	2220
QY	2220	taaacgcaggcgtgcgcacagtcacacactactgaatatcaattgatgctaagaatg	2279
DB	2221	TAAACGCAAGGCGCTGGCATCAGTGTCAACCTACTGAAATATCAGTTGATGCTAAGAGTG	2280
QY	2280	gcaatgttacccgccccaaacttaacaattggcgtgaaacaccaccagctttaacagtgatg	2339
DB	2281	GCAATGTACCGCCCCAACTTACAACATTGGCGGTGAANACCACCGNGCTTACAGTGATG	2340
QY	2340	gcactagtgtataatttagtgttaagggtagtggtgcagacaatatagcttagttaccgcgcg	2399
DB	2341	GCACGTAGTATAAATTTAGTGTTAAGGTAGTGSTACGAAACAATAAGCTTAGTTACCGCGC	2400
QY	2400	aacatttggcaagctactcaatgaagtcgaatcgaacgcgctgacagtgctctacaaaagt	2459
DB	2401	AACATTTTGGCAAGCTATCTAAATCAAGTCAATCGAAGCGCTGACAGTGCTCTACAAGACT	2460
QY	2460	ttaccgttaaagaagaagacgatgatagcgcgaacgctataccctggcctaaagatacga	2519
DB	2461	TTACCGTTTAAGAGAGACGATGATGACGCCAACGCTATCACCGTGGCTTAAGATACGA	2520
QY	2520	caaaaaatgccggcgcgagtcagcatcttaaaactcaaggtgtaaaacggtctaacgggtg	2579
DB	2521	CAAAAAATGCCGGCGCAGTCAGCATCTTAAACTCAAAAGGTAAAAACGGTCTTAACGGTGTG	2580
QY	2580	ctaccaaaaagatggttacgggttacctttgggcttagccaaagatagcggtctgacccattg	2639
DB	2581	CTACCCAAAAAGATGGTACCGGTTTACCTTTTGGGCTTAGCCCAAGATAGCGGTCTGACCATG	2640
QY	2640	gcaaaagcaccttaaacacacgatggcttgactgttaaagataccacgaacacaaatccaaag	2699
DB	2641	GCANAGCACCTTAACAACACGATGGCTTGACTGTTTAAAGATACCAACGACNAATATCCAAG	2700
QY	2700	tcggtgctaattgccaataattacttaattgtaattgtaattgcagtaaccaggtactggcattg	2759
DB	2701	TCGGTGCCTAAATGGCATTTAAATTTACTAATGTGAATGGTAGTAATCAGGTACTGSCATTG	2760
QY	2760	caaataccgcttcgcatcaccagagataaaattggcttttctggttctctgatggtcgagttg	2819
DB	2761	CAAAATACCGCTTCGCATTTACCAGAGATAAAATTTGGCTTTGCTGGTTCTGATGGTCAGTTG	2820
QY	2820	atacaaacaaaccttacttgatcaagacaagctacaagtggcgaattgaagtattacca	2879
DB	2821	ATACAAACAAACCTTATCTTGATCAAGCAAGCTACAAGTTGGCAATGTTTAGATTACCA	2880
QY	2880	acactggcatttaacgcagggtggtaaagccatacaggggctgtcccaacactgcctagca	2939
DB	2881	ACACTGGCATTTAACGCAGGTGGTAAAGCCATCATACGGGCTGTCCCCAACACTGCCTACCA	2940
QY	2940	ttgcgatcaaaagttagcgcgcaacatagaaactggggcaatacaatccaaagacaagaacaat	2999
DB	2941	TTGGCGGATCAAGGTAGCGCGAACATAGAACTGGGGCAATACAATCCAAGACAAGACAAT	3000
QY	3000	ccaacgctgcgacattaatgatataataatacagcgtttaacctaaaaataataaca	3059
DB	3001	CAACGCTGCCAGCATTAATGATATATTAATAACAGGCTTTAACCTTAAAAATAATAACA	3060
QY	3060	accctattgactttgtctccacttatgacattgtgactttgccaatggcaatggcaccaca	3119
DB	3061	ACCCATTGACTTTTGCTCCACTTATGACATTGTTGACTTTGCCAATGCAATGCCACCA	3120
QY	3120	ccgccaagttaaccatgataccgcgttaacaaaaccagtaagtgggtatagtgtgaatg	3179
DB	3121	CGCGCACAGTAAACCATGATACCGCTTAACAAAAACCAAGTAAAGTGGTATATGATGTGAATG	3180
QY	3180	tggatgataaacaccattctctaacaggcactgatacaataaaaaaactggctcctaaa	3239
DB	3181	TGGATGATACAAACCATTCATCTAACAGGCACTGATGACAAATAAAAAACTTGGCGTCAAAA	3240
QY	3240	ccaccaaactgaacaaacaaagtcctaattgtaatacagcaagctaaacttaattgataact	3299

|||||  
Db 3241 CCACAACTGACAAACAAAGTCTTAATGGTAATACAGCAACTAACTTAAATGTTAACT 3300  
QY 3300 ctatgtatgaatgaccccttgaatgaccccaagacatgcgcaaaatctaaacacccctag 3359  
Db 3301 CTAGTATGAATGACCTTGTAAAGCCAAAGACATCGCCGAAATCTAAACACCCCTAG 3360  
QY 3360 ccaaggaaattcacacacacaaagcacagacagacaccccttacaacaccttaccgtta 3419  
Db 3361 CCAAGGAAATTCACACACCAAGGACACAGACACACCGCCCTACAAACCTTTACCGCTTA 3420  
QY 3420 aaaaggtagatgaataataatgctgtagaccccaacccatcacccctgggtcaaaaga 3479  
Db 3421 AAAAGGTAGATGAATAATATGCTGATGACGCCAACGCCATCACCGGTGGGTCAAAGA 3480  
QY 3480 acgcaataataagtaacacacacccactcaaaagtgtaaaacggtcttaataataaa 3539  
Db 3481 ACGCAATAATCAAGTCACACCCCTAACACTCAAGGTGAACCGGTCTTAATATTAA 3540  
QY 3540 ccgacaaaatggtacggttacctttggcatatacaccaacacggtctttaaagccggca 3599  
Db 3541 CCGCAAAATGGTACGGTTACCTTTGGCATTAAACACCAACCGGTCTTAAAGCCGGCA 3600  
QY 3600 aaagcacccctaaagcaggtgggtgtctatataaaacccactgtagcgaacaaatcc 3659  
Db 3601 AAAGCACCTTAACGACCGGTGGTGTCTATTAAACACCCACTGGTAGCAACAAATCC 3660  
QY 3660 aagtcggtgtagtggtgaagttgcgaagtttaataataatggtgtgtagtgctg 3719  
Db 3661 AAGTCGCTGTGATGGCGTGAAGTTGCCAAGGTTAATAATATGTTGTGTAGTGCTG 3720  
QY 3720 gcatgtatggcacacacgattaccagagatgaattggctttactgggactaatggct 3779  
Db 3721 GCATTGTGCGACAACTCGCATTACCAGATGAATTTGGCTTTACTTGGGACTTAATGGCT 3780  
QY 3780 cactgtataaagaacacccaccctaaagcaagacgacatgaacgaggtggttaaaaga 3839  
Db 3781 CACTTGTAAAGCAACCCACCCTAAGCAAGACCGCATTAACGACGAGTGGTAAAGAA 3840  
QY 3840 ttacacacattcaatcaggtgagattgcccacacacgcatgctgtgacagggcgca 3899  
Db 3841 TTACCAACTTCAATAGGTGAGATTGCCCAACACCCATGATGCTGTACAGCGGCA 3900  
QY 3900 agattatgatttaaaacogaaactgaaacacaaaaacagcagtagtgcacaaacagcac 3959  
Db 3901 AGATTATGATTAAACCCGAATTTGAAACAAAAATCAGCAGTACTGCCAAAACAGCAC 3960  
QY 3960 aaactcattacagaaattcagtagcagatgaacaggttaataactttacggttagta 4019  
Db 3961 AAACCTCATTAACGAATTCACGTAGCAGATGAACAAAGGTAACTTTACGGTTAGTA 4020  
QY 4020 acccttactccagttatgacacccctcaagacccctctgtagtcaaccccttgcaggtgaaa 4079  
Db 4021 ACCCTTACTCCAGTTATGACACCTCAAGACCTCTGATGTATCATCCCTTTGCAGGTGAAA 4080  
QY 4080 acggtattacacacaaagtaataaaggtggtgctggtggtggtggtggtggtggtggtggt 4139  
Db 4081 ACGGCATTTACCAACCAAGTAAATAAAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 4140  
QY 4140 gcttaacacacccctaaagtcaggtgggtggtggtggtggtggtggtggtggtggtggtggt 4199  
Db 4141 GCTTAACACACGCTTAAGCTGACCGGTGGTGAATAATAATGCAAAAGGCAATTTGTCATTGACA 4200  
QY 4200 gccaaaaatggtaataataccatcacagagatgaacacacactctgtagttaaattaccacatg 4259  
Db 4201 GCCAAAATGGTCAAAATACCATCAGGACTAAGCAACACTCTAGCTAATGTTTACCAGTG 4260  
QY 4260 ataaaggttaggtacgcacacagaaagggcaataataataaagacaggaagacacaccc 4319  
Db 4261 ATAAAGGTAGGTAGCGTACGACACACAGAACAGGCGCAATATAATCAAGACGACACAAACCC 4320  
QY 4320 tgcgcgcacagcatgtgtgatgtgtaagcgcaggtcttaacttgcagggcaatggtaag 4379  
|||||

Db 4321 GTGCCGCCAGCATTTGTGATGTGCTAAGCGCAGGCTTTAACTTGCAAAGGCAANTGGTGAG 4380  
QY 4380 cgggtgacttgcctccacttatgacacccgtcaactttgcgagtggaatgccaccacccg 4439  
Db 4381 CGGTTCACTTTGCTCCACTTATGACACCGTCAACTTTGCCGATGGCAATGCCACCACCG 4440  
QY 4440 ctaaggtgacctatgtagacacaaagcaaaacagtaaaagttggttctatgtagtcaatgtg 4499  
Db 4441 CTAAGGTGACCTATGATGACACAAAGCAACAGCTAAAGTGTCTTATGATGTCATGTTGG 4500  
QY 4500 atgatacaacacattgaagttaaaagataaaaaacttgcgttaaaacccaccacattgacca 4559  
Db 4501 ATGATACAAACATTGAAGTTAAAGATAAAACTTTGGCGTTAAACACCAACCATTTGACCA 4560  
QY 4560 gtaaggcacaggtgctataaattggcctaaagcaatacaagctactggcgatgcgctg 4619  
Db 4561 GTACTGGCACAGTCTTAATAATTTGCCCTAAGCAATCAAGCTACTGGCGATGGCTTG 4620  
QY 4620 tcaaggcaagtagatcgttgcctataaacacacacttatctggagacatccaaactgcca 4679  
Db 4621 TCAAGGCCAGTGTATGCTTGCCTTAACACACCTTATCTGGCGACATTCCAAACTGCCA 4680  
QY 4680 aaggggcaagcgaacgaacactcagcggctaagtgtgtagtgcgtgagcaataaaggtca 4739  
Db 4681 AAGGGCAAGCCCAAGCAACTCAGCAGGCTATGTGATGCTGATGCAATAAGGTCA 4740  
QY 4740 tctatgacagtaccgataaacaagtactatcaagccaaatgatggcagctgtaaaaa 4799  
Db 4741 TCTATCACAGTACCAGTAAACAAAGTACTTATCAAGCCAAATAATGATGGCACAGTTGATAAAA 4800  
QY 4800 ccaagaagttgccaaagacaaactggcgcccaagcccaacccacagatgcacatlgg 4859  
Db 4801 CCAAGAAAGTTGCCAAAGCAAACTGGTCGCCCAAGCCCAACCCAGATGCCACATTTGG 4860  
QY 4860 ctcaaatgtaatgcaaatcagtcattaaacaaagaaacagtaaatgatgcataataaaagc 4919  
Db 4861 CTCAAATGATGTCAAATCAGTCATTAAACAAAGAACAGTAATGATGCCAATAAAAGC 4920  
QY 4920 aaggatcaatgaagacacgcctttgttaaaggactgaaaaagccgcttctgataaaca 4979  
Db 4921 AAGGCATCAATGAAGCAACCGCTTTGTGTTAAAGGACTTGAAGAGCGCTTCTGATAACA 4980  
QY 4980 aaacaaaaacgcgcagtaactgltgggtgatttaaatgcccgttgcacacacccgctga 5039  
Db 4981 AAACCAAAACCGCGCAGTAACGTGTGGGTGATTTAAATGCCGTGGCCCAACACCGCTCA 5040  
QY 5040 cctttgcaggggatacagcgaacacggcttaaaaaactggcgagacatttgacacacaaag 5099  
Db 5041 CCTTTCAGGGGATACAGGCACACCGCTTAAACAACTGGGCGAGACTTTGACCATCAAG 5100  
QY 5100 gtgggcaaacagacacaaataagctaaacgataaataacatcggtggtgtagcaggtactg 5159  
Db 5101 GTGGGCAACAGACACCAATAAGCTAACCGATAATAACATCGGTGTGTAGCAGGTACTG 5160  
QY 5160 atggcttcaactgtcaaaacttgcacaaagacctaaccacatcttaacacgcttaagtgaagtg 5219  
Db 5161 ATGGCTTCACTGTCAAACTTGCACAAAGACCTTAACCAATCTTAACACCGTTAATGCAAGTG 5220  
QY 5220 gcaacaaaattgatgacaaagcgctgtcttttgaactcaagcgtcaagcccaagcaaa 5279  
Db 5221 GCACCAAAATGTATGACAAAGGCGTGTCTTTGTAGACTCAAGCGGTCAAGCCCAAGCA 5280  
QY 5280 acacccctgtgctaaagtgcacaaaggcctggacccctgggtggcaaggtctaatgtaagtgg 5339  
Db 5281 ACACCCCTGTGCTAAGTGGCCAATGGCTGGACCTGGGTGGCAAGGTCAATCAGTAATGTGG 5340  
QY 5340 gcaaaagcaaaaagataccacgacgctgcacaaatgtatacaacagttaaacgaagtaacgaact 5399  
Db 5341 GCAAGGCACAAAAGATPACCGACGCTGCCAATGTAGAACAGTTTAAACAGAGTACGCAACT 5400  
QY 5400 tgttgggtctgtgaatgctggtaataacgctgacgcaatcaggttaaacattgccc 5459  
Db 5401 TGTGGGTCTTGGTAAATGCTGGTAATGATAACGCTGACGGCAATCAGGTAAACATTGCCG 5460  
|||||

Qy	5460	acatcaaaaaagcccaaaattcagggttcattcatctaatcccgcactgtcatcaaaagcaggca	5519
Db	5461	ACATCAAAAAGACCCAAATTCAGGTTCATCATCTAAACCGCACTGTCTATCAAAAGCAGGCA	5520
Qy	5520	cgtactctgcggttaaggtaataacagataccgaaaaacttgcactgggtggtatatacaag	5579
Db	5521	CGGTACTTGGCGGTAAAGGTAATAACGATACCGAAAAACTTGGCACCTGGTGTATACAAG	5580
Qy	5580	tgggcgtgataaaagacgcgaacgcgtaacggcgatttaagcaatgtttgggtcaaaaccc	5639
Db	5581	TGGGCGTGGATAAAGACGCGAACGCTAACGGCGATTTAAGCAATGTTTGGTCAAAACCC	5640
Qy	5640	aaaaagatggcagcaaaaaagccctgctcgcacattataacgcgcgaggttcagacaaact	5699
Db	5641	AAAAAGATGGCAGCAAAAAGCCCTGCTCGCGACCTTATAACGCCGCGAGGTCAGACCAACT	5700
Qy	5700	atttgaccacaacccccgaagaagcattgacagataataatgaacaaggtatccgcttct	5759
Db	5701	ATTITGACCAACCCCGCAGAAGCCATTGACAGAAATAATGAACAAGGTATCCGCTTCT	5760
Qy	5760	tccatgtcaacgatggcaatcaagacctgttgtaacaaggcgtaacgcgcatlgactcaa	5819
Db	5761	TCCATGTCAACGATGGCAATCAAGAGCCTGTGGTACAAAGGCGTAAACGGCATTTGACTCAA	5820
Qy	5820	gtcctcagcgaagcaactcagtggcgataggtttccaggccaaaggcagatggttgaagcog	5879
Db	5821	GTGCCTCAGGCAGACACTCAGTGGCGATAGGTTTCCAGGCCAAGGCAGATGGTGAAGCCG	5880
Qy	5880	ccgttgccataggcagaaaaccccaagcagcgaaccaactccatccgcatcggtgataaag	5939
Db	5881	CCGTTGCCATAGCAGACAAACCCAAAGCAGCAACCAATCATCCCATCGGTATTAACG	5940
Qy	5940	cacaaagccagggcgatcaatccatcgccatcggtacaggcaatgtgtagcaggttaagc	5999
Db	5941	CACAAGCCAGGGCGATCAATCCATCGCATCGGTACAGGCAATGTGGTAGCAGGTAAGC	6000
Qy	6000	actctggtgccatcggcgagcccaagcaactgtttaaggctgtataacagttacagttgctgggt	6059
Db	6001	ACTCTGTGCCATCGCGCACCCAAAGCACTGTTTAAGGTGTATTAACAGTTACAGTGTGGGTA	6060
Qy	6060	atacaaacaggtttaccgatgcactcaaaccgatgtctttggtgtgggccaataacatca	6119
Db	6061	ATAACAACCAAGTTTACCGATGCCATCAAAACCGATGCTTTTGGTGTGGGCAATAACATCA	6120
Qy	6120	ccgtgaccgaagaactcggttgccttaggttcaaaacttgcocatcagtcgaggcacac	6179
Db	6121	CCGTGACGAAAGTAATCGGTTGCTTTAGGTTCAAACCTCTGCCATCAGTGCAGGCACAC	6180
Qy	6180	acgcaggcacacaagcccaaaaaatctgcgcgcacagcaggtataaacaccacagcagggtg	6239
Db	6181	ACGCAAGCACACAAAGCCAAANAATCTGACGCGCACAGCAGGTACAAACCCACCACAGCAGGTG	6240
Qy	6240	caaccggtaacggttaaaaggcttcttgacaaaagcggttgggtgcggtctccggtgggtg	6299
Db	6241	CAACCGGTACGGTTAAAGCGTTTGTGGACAAACCGCGGTTGGTTCGCGTCTCCGTGGGTG	6300
Qy	6300	cctcaggtgctgaacgcgataccaaaaatgtgacagcaggttgaggtcagttgcaccagca	6359
Db	6301	CCTCAGGTGCTGAACCCGCTATCCAAAATGTGGCAGCAGGTGAGGTGAGTGCACACGCA	6360
Qy	6360	ccgatcggtcgaatggttagcgaattgtcaaaagccaccacaagcatgtccaaacycaacca	6419
Db	6361	CCGATGCGGTCAATGTGTAGCCAGTTGTCAAAGCCACCACCAAGGATTTGCCAACGCAACCA	6420
Qy	6420	atgagcttgaccatcgataccaccaaaacgaaaaataaggccaatgycagggtattcatcag	6479
Db	6421	ATGAGCTTGACCATCGTATCCACCAAAACGAAAAATAAGGCCCAATGCAGGGATTTTCATCAG	6480
Qy	6480	cgatggcgatggcggtccatgcccacaagcctacattccttggcagatccatggttacccggg	6539
Db	6481	CGATGGCGATGGCGTCCATGCCACAAGCGCTACATTCCTGGCAGATCCATGGTACCGGGG	6540

## RESULT 6

US-08-945-567A-1

US 08945567A

: GENERAL INFORMATION:

; APPLICANT: SASAKI, Ken

APPLICANT: HARKNESS, Robin E.

APPLICANT: LOOSMORE, Sheena M.

APPLICANT: CHONG, Pele

APPLICANT: KLEIN, Michel H

;	TITLE OF INVENTION:	HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF
;	TITLE OF INVENTION:	MORAXELLA

;; FILE REFERENCE: 1038-745 MIS

; CURRENT APPLICATION NUMBER: US/08/945,567A

; CURRENT FILING DATE: 1998-03-19

; PRIOR APPLICATION NUMBER: 08/431,718

; PRIOR FILING DATE: 1995-05-01

; PRIOR APPLICATION NUMBER: 08/478,370

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: 08/621,944

PRIOR FILING DATE: 1996-03-26

; PRIOR APPLICATION NUMBER: PCT/CA96/00264

; PRIOR FILING DATE: 1996-04-29

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn Ver. 2.1

;	SEQ	ID	NO	1
---	-----	----	----	---

; LENGTH: 6973

; TYPE: DNA

ORGANISM: Moraxella catarrhalis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (708)..(6683)

US-08-945-567A-1

### Query Match

Query Match 33.0%, Score 6301, DB 13, Length 6373,  
Best Local Similarity 100.0%: Pred No 0:

Best Local Similarity 100.0%, Fied: NO: 0:  
Matches 6972: Conservative 0: Mismatches

1 ccattggatatgggcagggtgtactcgccctgccgtatgatggcgatgacacccatttgccc 60

|||

Db 1 coactgatatggcaggtgtgtctgcgcgtatgtatggcgatgacacccatttgccc 60  
Qy 61 catactgacagattgacatgtgatattgaatgaatgacatgatttaacattgttt 120  
Db 61 catactgacagattgacatgtgatattgaatgaatgacatgatttaacattgttt 120  
Qy 121 aatactgttgccatcattaccataatttagtaaacgatttagtaacgcatattgtaaaat 180  
Db 121 aatactgttgccatcattaccataatttagtaaacgatttagtaacgcatattgtaaaat 180  
Qy 181 catgagccctttatgtgatcatatagaatagaatattatgattgattgatttattgt 240  
Db 181 catgagccctttatgtgatcatatagaatagaatattatgattgattgatttattgt 240  
Qy 241 atcagaatgggtgatgtatattgattgattgattgattgattgattgattgattgattg 300  
Db 241 atcagaatgggtgatgtatattgattgattgattgattgattgattgattgattgattg 300  
Qy 301 atttgatattttgaaactaatctattgattgattgattgattgattgattgattgattg 360  
Db 301 atttgatattttgaaactaatctattgattgattgattgattgattgattgattgattg 360  
Qy 361 taatgttaggtttttgaaataacacatgcaatattgttctactgttactacatgct 420  
Db 361 taatgttaggtttttgaaataacacatgcaatattgttctactgttactacatgct 420  
Qy 421 tgaatgacgatcccaatcaccagattcattcaagtgatgtgtttgtatcacgacatttta 480  
Db 421 tgaatgacgatcccaatcaccagattcattcaagtgatgtgtttgtatcacgacatttta 480  
Qy 481 cccataattatttcaataatgcctatgtatgcagtgatgcatttttaagtgaaacac 540  
Db 481 cccataattatttcaataatgcctatgtatgcagtgatgcatttttaagtgaaacac 540  
Qy 541 catgaatcacatctataaagtcatctttaaacaagccacagggcacatttatggcagtgcc 600  
Db 541 catgaatcacatctataaagtcatctttaaacaagccacagggcacatttatggcagtgcc 600  
Qy 601 agaglacgccaatcccaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 659  
Db 601 agaglacgccaatcccaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 660  
Qy 660 gtgtatgacactgagctttgcccgatttgcgcgctgcgtgcgtgcgtgcgtgcgtgcgtg 719  
Db 661 gtgtatgacactgagctttgcccgatttgcgcgctgcgtgcgtgcgtgcgtgcgtgcgtg 720  
Qy 720 cgtcagtgagcagcttctatgtctcaaaaaaagataaccaacatatacgaatttggaac 779  
Db 721 cgtcagtgagcagcttctatgtctcaaaaaaagataaccaacatatacgaatttggaac 780  
Qy 780 aaaaacagccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 839  
Db 781 aaaaacagccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 840  
Qy 840 gtgaaatgtcaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 899  
Db 841 gtgaaatgtcaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 900  
Qy 900 tcaatlggaagcagcttggataagataggttacogattgctacgggtcaagagtcacatcgcca 959  
Db 901 tcaatlggaagcagcttggataagataggttacogattgctacgggtcaagagtcacatcgcca 960  
Qy 960 tgggtgggtgataaagcagctgattgattgattgattgattgattgattgattgattgattg 1019  
Db 961 tgggtgggtgataaagcagctgattgattgattgattgattgattgattgattgattgattg 1020  
Qy 1020 atttgcctgacagcagctgattgattgattgattgattgattgattgattgattgattgattg 1079  
Db 1021 atttgcctgacagcagctgattgattgattgattgattgattgattgattgattgattgattg 1080  
Qy 1080 ttaacggccatgagctatttaaaagaaatcacgaagctcaaaagataatgatgtataaata 1139  
Db 1081 ttaacggccatgagctatttaaaagaaatcacgaagctcaaaagataatgatgtataaata 1140

Qy 1140 gacgcacacacgcaagcggacacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 1199  
Db 1141 gacgcacacacgcaagcggacacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 1200  
Qy 1200 gtcatttttccaaacgcttttggtacacgaggaacgacgacgacgacgacgacgacgacgacgacg 1259  
Db 1201 gtcatttttccaaacgcttttggtacacgaggaacgacgacgacgacgacgacgacgacgacgacg 1260  
Qy 1260 tgggtcttgccgacacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 1319  
Db 1261 tgggtcttgccgacacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 1320  
Qy 1320 ctactgcttgggagcgatagccttgggtgacgaggttactcgtgctcagctcagctcagctcagctcag 1379  
Db 1321 ctactgcttgggagcgatagccttgggtgacgaggttactcgtgctcagctcagctcagctcagctcag 1380  
Qy 1380 ttgcoctagggtcaaggttctgttgcactcagagtgataataattctagacgagcagctata 1439  
Db 1381 ttgcoctagggtcaaggttctgttgcactcagagtgataataattctagacgagcagctata 1440  
Qy 1440 caccacaaatcaccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1499  
Db 1441 caccacaaatcaccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1500  
Qy 1500 cactttccatttgggtgatttactctatacacaagcgttaaaatcactcactcactcactcactcact 1559  
Db 1501 cactttccatttgggtgatttactctatacacaagcgttaaaatcactcactcactcactcactcact 1560  
Qy 1560 ataaaacggatgcggtcacttgcacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1619  
Db 1561 ataaaacggatgcggtcacttgcacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1620  
Qy 1620 gtgaaattacttttcaggggtgatataacagctactgcagtgataaaatagatttggataata 1679  
Db 1621 gtgaaattacttttcaggggtgatataacagctactgcagtgataaaatagatttggataata 1680  
Qy 1680 ctttaactattaaaggttgggtgcagagacacacgatttaacccgataataataatcggtgtggtg 1739  
Db 1681 ctttaactattaaaggttgggtgcagagacacacgatttaacccgataataataatcggtgtggtg 1740  
Qy 1740 taaaagagcgtgataatagtggtctgaaagttaaacttgcataaaacttttaaaacacattcta 1799  
Db 1741 taaaagagcgtgataatagtggtctgaaagttaaacttgcataaaacttttaaaacacattcta 1800  
Qy 1800 ctgaggtgataacacactacattaaatgcacacacacacacacacacacacacacacacacacacac 1859  
Db 1801 ctgaggtgataacacactacattaaatgcacacacacacacacacacacacacacacacacacacac 1860  
Qy 1860 gtactacagcgtgaattattgattgattgattgattgattgattgattgattgattgattgattgattg 1919  
Db 1861 gtactacagcgtgaattattgattgattgattgattgattgattgattgattgattgattgattgattg 1920  
Qy 1920 aagcacaagcaaaaacccgtctatggtttaaagggtgaaagttaaacttgcataaaacttttaaaacac 1979  
Db 1921 aagcacaagcaaaaacccgtctatggtttaaagggtgaaagttaaacttgcataaaacttttaaaacac 1980  
Qy 1980 caacagcagcaatcgccactactcgtattaccagagataaaaattgggttgcctgcagagctg 2039  
Db 1981 caacagcagcaatcgccactactcgtattaccagagataaaaattgggttgcctgcagagctg 2040  
Qy 2040 gtgattgtgataaataaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2099  
Db 2041 gtgattgtgataaataaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2100  
Qy 2100 ttgcaattaccatagacaaatggcattgagcaggtgataaataaagcagcagcagcagcagcagcagc 2159  
Db 2101 ttgcaattaccatagacaaatggcattgagcaggtgataaataaagcagcagcagcagcagcagcagc 2160  
Qy 2160 aaggtagcagtgctaacagatgcggttaccatgcgaacagctcaaaagcagcagcagcagcagcagcagc 2219  
Db 2161 aaggtagcagtgctaacagatgcggttaccatgcgaacagctcaaaagcagcagcagcagcagcagcagc 2220

[illegible]

3301	Db	 ctagt	gatgaagatgccccttg	taagccgaagacatcgcgaaaa	lctaaacaccc	tag	3360											
3360	QY	ccaag	gaattcacaccacaag	gcacagacacccgcctacaac	ctttacc	gtta	3419											
3361	Db	ccaag	gaattcacaccacaag	gcacagacacccgcctacaac	ctttacc	gtta	3420											
3420	QY	aaaag	tagatgaaaataa	tactgtgatgacgccaa	gcgcataccacgtgggt	cataaaga	3479											
3421	Db	aaaag	tagatgaaaataa	tactgtgatgacgccaa	gcgcataccacgtgggt	cataaaga	3480											
3480	QY	acgca	aatatcaag	tcaacacccctaac	ctcaaaagtgaaaa	cgggtcttaattaaaa	3539											
3481	Db	acgca	aatatcaagtcaac	acccctaacctcaag	tgaaaacggtcttaatt	aaaa	3540											
3540	QY	ccgaca	aaaaatgtag	cgggttaccttggcatt	taacaccacaagcgtctta	aacgcgcga	3599											
3541	Db	ccgaca	aaaaatgtagcgggt	taccttggcatttaac	accacaagcggctta	aacgcgcga	3600											
3600	QY	aaag	caccttaaacgac	gggttgctatataaa	ccccccactggtag	cgacaacatcc	3659											
3601	Db	aaag	caccttaaacgac	gggttgctatataaa	ccccccactggtag	cgacaacatcc	3660											
3660	QY	aagt	cgggtcgtgat	ggcgtgaagttgcc	aaggttaataatgggt	tgttagtgctg	3719											
3661	Db	aagt	cgggtcgtgat	ggcgtgaagttgcc	aaggttaataatgggt	tgttagtgctg	3720											
3720	QY	gcatt	gatgcacaactc	gcattaccagacgat	gaattggctttact	gggaactaatgct	3779											
3721	Db	gcatt	gatgcacaactc	gcattaccagacgat	gaattggctttact	gggaactaatgct	3780											
3780	QY	cact	gtataaaagca	ccccccactaag	caaaagcgcattaa	cgcagcgggtgg	taaaaga	3839										
3781	Db	cact	gtataaaagca	ccccccactaag	caaaagcgcattaa	cgcagcgggtgg	taaaaga	3840										
3840	QY	ttacca	aacttcaatcag	gtgagattgcc	caaaacagccat	gatgctgctgac	aggcgcga	3899										
3841	Db	ttacca	aacttcaatcag	gtgagattgcc	caaaacagccat	gatgctgctgac	aggcgcga	3900										
3900	QY	agatt	tattgattaaaa	accggaacttgaa	acaaaatcag	cagctactgc	caaaacagcac	3959										
3901	Db	agatt	tattgattaaaa	accggaacttgaa	acaaaatcag	cagctactgc	caaaacagcac	3960										
3960	QY	aaaac	tcatcac	gaattctcag	tagcagatgaaca	agggtataac	actttac	ggttagta	4019									
3961	Db	aaaac	tcatcacgaatt	ctcagtagcagat	gaagaaggtata	acactttac	ggttagta	4020										
4020	QY	accct	tactccagt	tatgacac	ctcaagac	ctctgtatgc	atcac	ctttgcagg	tgaaa	4079								
4021	Db	accct	tactccagt	tatgacac	ctcaagac	ctctgtatgc	atcac	ctttgcagg	tgaaa	4080								
4080	QY	acgg	cattacc	caaggtataa	taaaaggtgtg	gtgcgtgggc	catgc	acaaac	ccaaag	4139								
4081	Db	acgg	cattacc	caaggtataa	taaaaggtgtg	gtgcgtgggc	catgc	acaaac	ccaaag	4140								
4140	QY	gctt	aac	ccgcgt	taagctgac	gcgtgggtta	ataatggc	aaagg	catgtcat	tgaca	4199							
4141	Db	gctt	aac	ccgcgt	taagctgac	gcgtgggtta	ataatggc	aaagg	catgtcat	tgaca	4200							
4200	QY	gccaaa	atggtc	aaaaat	taccat	cacagga	cttaag	caac	ctctag	cttaattg	caaatg	4259						
4201	Db	gccaaa	atggtc	aaaaat	taccat	cacagga	cttaag	caac	ctctag	cttaattg	caaatg	4260						
4260	QY	ataa	aggtag	cgtac	gcac	caagaa	caggg	caataata	taacaa	agacga	agacaaa	cccc	4319					
4261	Db	ataa	aggtag	cgtac	gcac	caagaa	caggg	caataata	taacaa	agacga	agacaaa	cccc	4320					
4320	QY	gtcc	ccgac	catgtt	gatg	tgc	taag	cgagg	ctt	taact	gtc	aggg	caatg	gaag	4379			
4321	Db	gtcc	ccgac	catgtt	gatg	tgc	taag	cgagg	ctt	taact	gtc	aggg	caatg	gaag	4380			
4380	QY	cgg	ttgact	ttg	ctc	act	tatg	ac	cgt	ta	act	ttg	ccg	atg	gc	acc	cgc	4439

Db 4381 cgggtgacttggctccacttatgaacacggtcaactttggcgaatggcaatgccaccacg 4440  
Qy 4440 ctaaggtgacctatgatgacacaaagcaaaacacgtaaaagtgggtctctatgatgtcaatgtg 4499  
Db 4441 ctaaggtgacctatgatgacacaaagcaaaacacgtaaaagtgggtctctatgatgtcaatgtg 4500  
Qy 4500 atgatacaacattgaagttaaaagataaaaaaacttggcgttaaaacaccaccatttgacca 4559  
Db 4501 atgatacaacattgaagttaaaagataaaaaaacttggcgttaaaacaccaccatttgacca 4560  
Qy 4560 gtactggcacaggtgctataaaattggcctaagcaatcaagctacttgccgatgcgcttg 4619  
Db 4561 gtactggcacaggtgctataaaattggcctaagcaatcaagctacttgccgatgcgcttg 4620  
Qy 4620 tcaagggcagtgatctgttgcctactataaacacttatctatctggtgcacatccaaactggca 4679  
Db 4621 tcaagggcagtgatctgttgcctactataaacacttatctatctggtgcacatccaaactggca 4680  
Qy 4680 aaggggcaagcgaagcaaacactcagcaggtctatgtggtgctgatggcaataaagggtca 4739  
Db 4681 aaggggcaagcgaagcgaaacactcagcaggtctatgtggtgctgatggcaataaagggtca 4740  
Qy 4740 tctatgacagtgatacaaaagtaactatcaagggcgaataatgatggcagttgataaaa 4799  
Db 4741 tctatgacagtgatacaaaagtaactatcaagggcgaataatgatggcagttgataaaa 4800  
Qy 4800 ccaaaagaattgccaagaagcaaacctgttcgcccgaagcccaaacccacagatggcacattgg 4859  
Db 4801 ccaaaagaattgccaagaagcaaacctgttcgcccgaagcccaaacccacagatggcacattgg 4860  
Qy 4860 ctcaaatgaatgctcaaatcagtcatttaacaaagaaacaaagtaaatgatggcgaataaaagc 4919  
Db 4861 ctcaaatgaatgctcaaatcagtcatttaacaaagaaacaaagtaaatgatggcgaataaaagc 4920  
Qy 4920 aagggcgaatgaagcaaacacgcttctgtttaaaggacttgaaagacgttgaaagcgttctgataaca 4979  
Db 4921 aagggcgaatgaagcaaacacgcttctgtttaaaggacttgaaagacgttctgataaca 4980  
Qy 4980 aaacaaaacgcgcagtaaacctgtgggtgatttaaatccgttggcccaaacacgcgtga 5039  
Db 4981 aaacaaaacgcgcagtaaacctgtgggtgatttaaatccgttggcccaaacacgcgtga 5040  
Qy 5040 ccttgcaggggatacaggcacaacacggttaaaaaactggggcagagactttgaccatcaaaag 5099  
Db 5041 ccttgcaggggatacaggcacaacacggttaaaaaactggggcagagactttgaccatcaaaag 5100  
Qy 5100 gtgggcaaacagacacacaaataagctaacacgataataacatcgtgtgtggtagcaggtactg 5159  
Db 5101 gtgggcaaacagacacacaaataagctaacacgataataacatcgtgtgtggtagcaggtactg 5160  
Qy 5160 atggcttcaactgtcaaaacttgcgaagacctaaccatacttaacagcgttaattgcaggtg 5219  
Db 5161 atggcttcaactgtcaaaacttgcgaagacctaaccatacttaacagcgttaattgcaggtg 5220  
Qy 5220 gcacaaaattgatgacaaagcgtgtcttctgactcaaacggttcaagccaaagcaa 5279  
Db 5221 gcacaaaattgatgacaaagcgtgtcttctgactcaaacggttcaagccaaagcaa 5280  
Qy 5280 acacccctgtctaaagtgcacaaatggcgtgacctgggtggcgaaggtcaatcagtaattgtgg 5339  
Db 5281 acacccctgtctaaagtgcacaaatggcgtgacctgggtggcgaaggtcaatcagtaattgtgg 5340  
Qy 5340 gcaagggcacaagatacaccagcgtcccaatgtacacaggttaaacgaagtacgcaact 5399  
Db 5341 gcaagggcacaagatacaccagcgtcccaatgtacacaggttaaacgaagtacgcaact 5400  
Qy 5400 tgttgggtctctggtaatgctggtaagtataacgctgacggaatcaatcagtaattgccc 5459  
Db 5401 tgttgggtctctggtaatgctggtaagtataacgctgacggaatcaatcagtaattgccc 5460  
Qy 5460 acatcaaaaaagaccccaattcaggttcaatcaatcaacgcactgtctcaaaagcagga 5519  
Db 5461 acatcaaaaaagaccccaattcaggttcaatcaatcaacgcactgtctcaaaagcagga 5520

Qy 5520 cggctacttggcggtaaaaggttaataacgataccgaaaaacttgcacatgggtgggtatacaag 5579  
Db 5521 cggctacttggcggtaaaaggttaataacgataccgaaaaacttgcacatgggtgggtatacaag 5580  
Qy 5580 tggcgctgggataaaagacggcgaacgctaacgcgatattgaagcaatgtttgggtcaaaaaccc 5639  
Db 5581 tggcgctgggataaaagacggcgaacgctaacgcgatattgaagcaatgtttgggtcaaaaaccc 5640  
Qy 5640 aaaaagatggcagcaaaaaaagcctgtcctcacttataacgcccaggtcagaccacact 5699  
Db 5641 aaaaagatggcagcaaaaaaagcctgtcctcacttataacgcccaggtcagaccacact 5700  
Qy 5700 atttgaccacaaccccgcagaaagccatttgacagataaaattgaacaagggtatccgcttct 5759  
Db 5701 atttgaccacaaccccgcagaaagccatttgacagataaaattgaacaagggtatccgcttct 5760  
Qy 5760 tccatgtcaacgataggcaaatcaagagcctgttggtaacagggtgaacggcattgactcaaa 5819  
Db 5761 tccatgtcaacgataggcaaatcaagagcctgttggtaacagggtgaacggcattgactcaaa 5820  
Qy 5820 gtgcctcaggaagcaactcagtgccgataagtttccagggccaagggcagatggtgaagccg 5879  
Db 5821 gtgcctcaggaagcaactcagtgccgataagtttccagggccaagggcagatggtgaagccg 5880  
Qy 5880 cggttgccatagggcagacaaaccccaagcaggcaacccaatccatcgcctatcgcctcgggtgataacg 5939  
Db 5881 cggttgccatagggcagacaaaccccaagcaggcaacccaatccatcgcctatcgcctcgggtgataacg 5940  
Qy 5940 cacaagcacggcgatcaatccatcgcctatcgcctcaggtacagggccaatgtgtgacaggtgaagc 5999  
Db 5941 cacaagcacggcgatcaatccatcgcctatcgcctcaggtacagggccaatgtgtgacaggtgaagc 6000  
Qy 6000 actctggtgccatcgcgcacccaagcactgtttaaaggtcgaataacagttcacagttcgggta 6059  
Db 6001 actctggtgccatcgcgcacccaagcactgtttaaaggtcgaataacagttcacagttcgggta 6060  
Qy 6060 ataaacacagtttaccgataccactcaaacgatacttcttgggtggtggcaataacatca 6119  
Db 6061 ataaacacagtttaccgataccactcaaacgatacttcttgggtggtggcaataacatca 6120  
Qy 6120 cggtagcgaagaaagtaactcggttgccttaggttcaaaactctgccatcagtgtaggacacac 6179  
Db 6121 cggtagcgaagaaagtaactcggttgccttaggttcaaaactctgccatcagtgtaggacacac 6180  
Qy 6180 acgagggcacacaagcccaaaaaatctgcagcgacacaggtacaaacccacacagcaggtg 6239  
Db 6181 acgagggcacacaagcccaaaaaatctgcagcgacacaggtacaaacccacacagcaggtg 6240  
Qy 6240 caacgggtacggttaaaagccttgcctgggacaaacggcggttgggtcctcctcgtgggtg 6299  
Db 6241 caacgggtacggttaaaagccttgcctgggacaaacggcggttgggtcctcctcgtgggtg 6300  
Qy 6300 cctcaggtgtgaaacgcgtatccaaaaatgtggcgaacaggtgaggtcagtcgaccacagca 6359  
Db 6301 cctcaggtgtgaaacgcgtatccaaaaatgtggcgaacaggtgaggtcagtcgaccacagca 6360  
Qy 6360 cggatcgggtcaatgttagccagtttatacaagcccaacaaagcattgccaaacgcaacca 6419  
Db 6361 cggatcgggtcaatgttagccagtttatacaagcccaacaaagcattgccaaacgcaacca 6420  
Qy 6420 atgagcttgacctcgtatcccaaaaaacgaaaaataaggccaatgagggatttctcag 6479  
Db 6421 atgagcttgacctcgtatcccaaaaaacgaaaaataaggccaatgagggatttctcag 6480  
Qy 6480 cgaatggcgtatggcgtccatgccacaagcctatacttctggcagatccatgggtaccgggg 6539  
Db 6481 cgaatggcgtatggcgtccatgccacaagcctatacttctggcagatccatgggtaccgggg 6540  
Qy 6540 gtattgccaccacaacgggtcaaggtgaggtgaggtggaggtctcgaaggtctcgagata 6599  
Db 6541 gtattgccaccacaacgggtcaaggtgaggtgaggtggaggtctcgaaggtctcgagata 6600







QY 1200 gtcaattttccaaagcctttgttacacgggcaacagctaaagatgcttattccttggcag 1259  
DB 1201 gtcaattttccaaagcctttgttacacgggcaacagctaaagatgcttattccttggcag 1260  
QY 1260 tgggtcttgccgccaagcagcgaggcccaatcaccatcgctattgttctgtagcaaat 1319  
DB 1261 tgggtcttgccgccaagcagcgaggcccaatcaccatcgctattgttctgtagcaaat 1320  
QY 1320 ctatgctgttggagcgatagcccttggctgcagggtactgtgctcagctacagggcagta 1379  
DB 1321 ctatgctgttggagcgatagcccttggctgcagggtactgtgctcagctacagggcagta 1380  
QY 1380 ttgcccctaggtcaaggttctgttgcacccagagtgagtaataattctagacggcctata 1439  
DB 1381 ttgcccctaggtcaaggttctgttgcacccagagtgagtaataattctagacggcctata 1440  
QY 1440 caccataatccaggcagctagaccccaagtttcaagccaccaataatacgaaggcggttc 1499  
DB 1441 caccataatccaggcagctagaccccaagtttcaagccaccaataatacgaaggcggttc 1500  
QY 1500 cactttccattggtagtaactctatacaacgttaaaatcataatgtcggctgaggtgtta 1559  
DB 1501 cactttccattggtagtaactctatacaacgttaaaatcataatgtcggctgaggtgtta 1560  
QY 1560 ataaaaccagatcggttcaatgttggcacagctagaagcggtgttgaagtggttcaaggagc 1619  
DB 1561 ataaaaccagatcggttcaatgttggcacagctagaagcggtgttgaagtggttcaaggagc 1620  
QY 1620 gtagaattacttttcagggttgatgataacagctactgacgtataaaaaataggtttggataata 1679  
DB 1621 gtagaattacttttcagggttgatgataacagctactgacgtataaaaaataggtttggataata 1680  
QY 1680 ctttaactattaaagtggtgcagagacaacagcattaaaccgataataatcgggtgtg 1739  
DB 1681 ctttaactattaaagtggtgcagagacaacagcattaaaccgataataatcgggtgtg 1740  
QY 1740 taaaagggctgataatagttggtctgaaagttaaacttgcataaaactttaacaacttta 1799  
DB 1741 taaaagggctgataatagttggtctgaaagttaaacttgcataaaactttaacaacttta 1800  
QY 1800 ctgaggtgaatacaactataataataatgccaacacagctaaagtgtagtagtagta 1859  
DB 1801 ctgaggtgaatacaactataataataatgccaacacagctaaagtgtagtagtagta 1860  
QY 1860 gtactacagctgaattattgagtgatgttttaacctttaccagcccaatacagggcagtc 1919  
DB 1861 gtactacagctgaattattgagtgatgttttaacctttaccagcccaatacagggcagtc 1920  
QY 1920 aaagcacaagcaaaacgctctaaggcgttaattgggttgaggtttactaataatgcagaaa 1979  
DB 1921 aaagcacaagcaaaacgctctaaggcgttaattgggttgaggtttactaataatgcagaaa 1980  
QY 1980 caacagcagcaatcgccactactcgtattaccagagataaaattgctttgctcgagatg 2039  
DB 1981 caacagcagcaatcgccactactcgtattaccagagataaaattgctttgctcgagatg 2040  
QY 2040 gtgatgttgatgaaacaagcaccatttttgataaaaaaacacttaaaagtggttagtg 2099  
DB 2041 gtgatgttgatgaaacaagcaccatttttgataaaaaaacacttaaaagtggttagtg 2100  
QY 2100 ttgcaattaccatagacaaatggcaatgtatgcagggttaataaaagatcagtaacttggcca 2159  
DB 2101 ttgcaattaccatagacaaatggcaatgtatgcagggttaataaaagatcagtaacttggcca 2160  
QY 2160 aaggtgacagtgctaaagatgagggttaccatcgaaacagctcaaaagccgcaagcctactt 2219  
DB 2161 aaggtgacagtgctaaagatgagggttaccatcgaaacagctcaaaagccgcaagcctactt 2220  
QY 2220 taaacgaggcgctggcactcagtgctacacactactgaaatcactcagttgagtgaagtg 2279  
DB 2221 taaacgaggcgctggcactcagtgctacacactactgaaatcactcagttgagtgaagtg 2280

QY 2280 gcaattgtaccgcccacttacacattgggtgaataccaccagcagcttaacagtgatg 2339  
DB 2281 gcaattgtaccgcccacttacacattgggtgaataccaccagcagcttaacagtgatg 2340  
QY 2340 gcaactagtgataaaatttagtttaagggtagtgggtacgaacaataagcttagttaccgcg 2399  
DB 2341 gcaactagtgataaaatttagtttaagggtagtgggtacgaacaataagcttagttaccgcg 2400  
QY 2400 aacatttggcagctctataatgaagtcaatcgaaacggctgacagtgctctacaaaagt 2459  
DB 2401 aacatttggcagctctataatgaagtcaatcgaaacggctgacagtgctctacaaaagt 2460  
QY 2460 ttaccgtttaaagaagaagacgatgatgcgcgaacgcttatcccgctggcctaagaatacga 2519  
DB 2461 ttaccgtttaaagaagaagacgatgatgcgcgaacgcttatcccgctggcctaagaatacga 2520  
QY 2520 caaaaaatgcggcgcgagtcagcatcttaaaactcaaaaggttaaaacggctctcaacgggtg 2579  
DB 2521 caaaaaatgcggcgcgagtcagcatcttaaaactcaaaaggttaaaacggctctcaacgggtg 2580  
QY 2580 ctacccaaaaagatggttacggtttacctttgggttagccaagatagcggtctgacccattg 2639  
DB 2581 ctacccaaaaagatggttacggtttacctttgggttagccaagatagcggtctgacccattg 2640  
QY 2640 gcaaaagcaccctaaacaaacagatgggtgactgtttaaagatacccaacgaacaaatcccaag 2699  
DB 2641 gcaaaagcaccctaaacaaacagatgggtgactgtttaaagatacccaacgaacaaatcccaag 2700  
QY 2700 tcgggtcctaagtgccattaaattacttaattgtaaggtagtaattccaggtactggcattg 2759  
DB 2701 tcgggtcctaagtgccattaaattacttaattgtaaggtagtaattccaggtactggcattg 2760  
QY 2760 caaataaccgctgcattaccagagataaaattgggttggctgtgtgtgtgtgtgtgtgtgtgt 2819  
DB 2761 caaataaccgctgcattaccagagataaaattgggttggctgtgtgtgtgtgtgtgtgtgtgt 2820  
QY 2820 atacaaaaaaacttacttctgatacagaacagctacaagttggcaatgttagattaccca 2879  
DB 2821 atacaaaaaaacttacttctgatacagaacagctacaagttggcaatgttagattaccca 2880  
QY 2880 acactggcattaaacgaggtggttaaaagccatcacagggtgtcccccaacactgcctagca 2939  
DB 2881 acactggcattaaacgaggtggttaaaagccatcacagggtgtcccccaacactgcctagca 2940  
QY 2940 ttgcgagataaaagtagccgcaacatagaaactgggccaatacaatccaaagcaaaagacaaat 2999  
DB 2941 ttgcgagataaaagtagccgcaacatagaaactgggccaatacaatccaaagcaaaagacaaat 3000  
QY 3000 ccaacgctgccagcattaatgatataataatacagggcttttaaccttaaaaaataataaaca 3059  
DB 3001 ccaacgctgccagcattaatgatataataatacagggcttttaaccttaaaaaataataaaca 3060  
QY 3060 acccaattgactttgtctccactttaagcattgtgtgactttgccaatggcaatggcaccaca 3119  
DB 3061 acccaattgactttgtctccactttaagcattgtgtgactttgccaatggcaatggcaccaca 3120  
QY 3120 cggccaagtaaccaatgatatacgcgttaacaaaacaggtaaagtggttatgtatgtgtaagtg 3179  
DB 3121 cggccaagtaaccaatgatatacgcgttaacaaaacaggtaaagtggttatgtatgtgtaagtg 3180  
QY 3180 tggatgatacaaacattcttaacagggcactgatgacaaataaaaaacttggcgctcaaaa 3239  
DB 3181 tggatgatacaaacattcttaacagggcactgatgacaaataaaaaacttggcgctcaaaa 3240  
QY 3240 ccaccaactgaacaaaacagtgctaaatgggttaatacagcaacttaactttaatgttaact 3299  
DB 3241 ccaccaactgaacaaaacagtgctaaatgggttaatacagcaacttaactttaatgttaact 3300  
QY 3300 ctatgataaagatgcccctgtttaaagcgaagacatcgccgaataatcttaaaacacccctag 3359  
DB 3301 ctatgataaagatgcccctgtttaaagcgaagacatcgccgaataatcttaaaacacccctag 3360  
QY 3360 ccaaggaataatcacaccacaaaggccacagcagacacccgctctacaaaactttaccggtta 3419



QY	5580	tgggcgtggataaagacgcaacgcgtaacgcgatttaagcaatgtttgggtcaaaaccc	5639
Db	5581	tgggcgtggataaagacgcaacgcgtaacgcgatttaagcaatgtttgggtcaaaaccc	5640
QY	5640	aaaaagatggcagcaaaaaagccctgctccacctataacgcgcgcaagtcagaccaact	5699
Db	5641	aaaaagatggcagcaaaaaagccctgctccacctataacgcgcgcaagtcagaccaact	5700
QY	5700	atttgacaaacaccccgagaaagccattgacagataaaatgaacaaggtatccgcttct	5759
Db	5701	atttgacaaacaccccgagaaagccattgacagataaaatgaacaaggtatccgcttct	5760
QY	5760	tccatgtccaatcgttgccaatacagagcgtgtgtacaaagggcgtaaacgcgattgactcaa	5819
Db	5761	tccatgtccaatcgttgccaatacagagcgtgtgtacaaagggcgtaaacgcgattgactcaa	5820
QY	5820	gtgcctcagcgaagcactcagtggcgtatagatttcccaggccaagcgagatggtgaagccg	5879
Db	5821	gtgcctcagcgaagcactcagtggcgtatagatttcccaggccaagcgagatggtgaagccg	5880
QY	5880	ccgttgccataggcagacaaccccaagcagggaaccaatccatcgccatcggtatataacg	5939
Db	5881	ccgttgccataggcagacaaccccaagcagggaaccaatccatcgccatcggtatataacg	5940
QY	5940	cacaagccacgggcgataatccatcgcacatcggtacaggcaatgtgtgtagcaggttaagc	5999
Db	5941	cacaagccacgggcgataatccatcgcacatcggtacaggcaatgtgtgtagcaggttaagc	6000
QY	6000	actctgtgtccatcggcgaccacaagcactgttaaggcgtgataaacagttacagtttggtgta	6059
Db	6001	actctgtgtccatcggcgaccacaagcactgttaaggcgtgataaacagttacagtttggtgta	6060
QY	6060	atacaaacacagtttacccgatgccatcaaacccgatgtcttgggtgtggcgaataacatca	6119
Db	6061	atacaaacacagtttacccgatgccatcaaacccgatgtcttgggtgtggcgaataacatca	6120
QY	6120	ccgtgaccgaagaactcgtgtgcttaggtgtcaaatctgccatccagtcaggcgacac	6179
Db	6121	ccgtgaccgaagaactcgtgtgcttaggtgtcaaatctgccatccagtcaggcgacac	6180
QY	6180	acgcagggcacaaagcccaaaaaatctgacgcgcacagcaggtacaaccaccacagcaggtg	6239
Db	6181	acgcagggcacaaagcccaaaaaatctgacgcgcacagcaggtacaaccaccacagcaggtg	6240
QY	6240	caaccgttcaggtttaaagcgttggcgcaaacgcggttgggtgcggtctccctcggtgggtg	6299
Db	6241	caaccgttcaggtttaaagcgttggcgcaaacgcggttgggtgcggtctccctcggtgggtg	6300
QY	6300	cctcaggtgtgaacgcgcgtatccaaaaatgtggcagcaggtgaggttcagtcgaccacaga	6359
Db	6301	cctcaggtgtgtgaacgcgcgtatccaaaaatgtggcagcaggtgaggttcagtcgaccacaga	6360
QY	6360	ccgatccggtccaatgttagccaggtgtacaagaaccaccaaaagcattgtccaacgcaacca	6419
Db	6361	ccgatccggtccaatgttagccaggtgtacaagaaccaccaaaagcattgtccaacgcaacca	6420
QY	6420	atgagcttgaccatcgtatcccaaaaaagaaataagggccaatgacggatttcacatcag	6479
Db	6421	atgagcttgaccatcgtatcccaaaaaagaaataagggccaatgacggatttcacatcag	6480
QY	6480	cgatggcgatggcggtccatgcacaaagcctcacattcctggcagatccatggttacccgggg	6539
Db	6481	cgatggcgatggcggtccatgcacaaagcctcacattcctggcagatccatggttacccgggg	6540
QY	6540	gtattggcaccaccaacaggttcaaggtgcggtggcagtgagggttcagagctgcgaagctgcggata	6599
Db	6541	gtattggcaccaccaacaggttcaaggtgcggtggcagtgagggttcagagctgcgaagctgcggata	6600
QY	6600	atggttcaatgggtattttaaatacaatgtgtcagccacatacccaagcccatgttagggggcgg	6659
Db	6601	atggttcaatgggtattttaaatacaatgtgtcagccacatacccaagcccatgttagggggcgg	6660

Qy	6660	cagttgtgtgcagggttttcaacttttaagccataaaatcgcaagatttttacttaaaaaatcaat	6711
Db	6661	cagttgtgtgcagggttttcaacttttaagccataaaatcgcaagatttttacttaaaaaatcaat	6720
Qy	6720	ctacacatagttgtatataaaacagcatcagcatcagtcacatatactatgatgctgatgttttt	6779
Db	6721	ctccacatagttgtataaaacagcatcagcatcagtcacatatactatgatgctgatgttttt	6780
Qy	6780	tataccttaaaacattttaccgcctcaagtgattctctttcaccatgaaccaaatacgccatt	6839
Db	6781	tataccttaaaacattttaccgcctcaagtgattctctttcaccatgaaccaaatacgccatt	6840
Qy	6840	gatactatggttaaaacttattgagtataatttttacaatgattgttttagatatggtctaaat	6899
Db	6841	gatactatggttaaaacttattgagtataatttttacaatgattgttttagatatggtctaaat	6900
Qy	6900	tgtgccattgaccaaataaagccagatttatcccgaaaatttctgattatgatcctgtga	6959
Db	6901	tgtgccattgaccaaataaagccagatttatcccgaaaatttctgattatgatcctgtga	6960
Qy	6960	cctgcaggtgcgac	6972
Db	6961	cctgcaggtgcgac	6973
RESULT 8			
US-08-945-567D-1			
; Sequence 1, Application US/08945567D			
; GENERAL INFORMATION:			
; APPLICANT: SASAKI, Ken			
; APPLICANT: HARKNESS, Robin E.			
; APPLICANT: LOOSMORE, Sheena M.			
; APPLICANT: CHONG, Pele			
; APPLICANT: KLEIN, Michel H.			
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN C			
; TITLE OF INVENTION: MORAXELLA			
; FILE REFERENCE: 1038-745 MIS			
; CURRENT APPLICATION NUMBER: US/08/945,567D			
; CURRENT FILING DATE: 1996-04-29			
; PRIOR APPLICATION NUMBER: 08/431,718			
; PRIOR FILING DATE: 1995-05-01			
; PRIOR APPLICATION NUMBER: 08/478,370			
; PRIOR FILING DATE: 1995-06-07			
; PRIOR APPLICATION NUMBER: 08/621,944			
; PRIOR FILING DATE: 1996-03-26			
; PRIOR APPLICATION NUMBER: PCT/CA96/00264			
; PRIOR FILING DATE: 1996-04-29			
; NUMBER OF SEQ ID NOS: 10			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 1			
; LENGTH: 6973			
; TYPE: DNA			
; ORGANISM: Moraxella catarrhalis			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (708)..(6683)			
US-08-945-567D-1			
Query Match 99.8%; Score 6961; DB 13; Length 6973;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 6972; Conservative 0; Mismatches 0; Indels 1; Gaps			
Qy	1	ccatgatatggcgaggtgtgctgcctgcctgctatgatgcgatgcacacccatttgcgc	60
Db	1	ccatgatatggcgaggtgtgctgcctgcctgctatgatgcgatgcacacccatttgcgc	60
Qy	61	catatctgtacgattttgaacatgtatatgatttaacatgtgacatgatatttaacattgttt	120
Db	61	catatctgtacgattttgaacatgtatatgatttaacatgtgacatgatatttaacattgttt	120
Qy	121	aatactgttgccatcatcaccataatttagtaaacgatttagtaaacgatttgaataat	180





Db 4501 atgatacaacacattgaagttaaagataaaaaaacttggcgttaaaaaacccacacattgacca 4560  
Qy 4560 gtactggcacagggtgctaataaaattgcccctaaagcaatacaagctactgagcgcttg 4619  
Db 4561 gtactggcacagggtgctaataaaattgcccctaaagcaatacaagctactgagcgcttg 4620  
Qy 4620 tcaaggccagtgatcgttctctatctataacaccttatcttgcgacatccaaactgcca 4679  
Db 4621 tcaaggccagtgatcgttctctatctataacaccttatcttgcgacatccaaactgcca 4680  
Qy 4680 aaggggcaagccaaagcgaacaaactcagcagctatgtggatgctgagtggaataagggtca 4739  
Db 4681 aaggggcaagccaaagcgaacaaactcagcagctatgtggatgctgagtggaataagggtca 4740  
Qy 4740 tctatgacagtaccgatacaacagtactatcaagccaaaaatgtaggacagttgataaaa 4799  
Db 4741 tctatgacagtaccgatacaacagtactatcaagccaaaaatgtaggacagttgataaaa 4800  
Qy 4800 ccaagaagaattgccaagacaaaactgctgcgccaaagcccaaacccacagatggcacattgg 4859  
Db 4801 ccaagaagaattgccaagacaaaactgctgcgccaaagcccaaacccacagatggcacattgg 4860  
Qy 4860 ctcaaatgaattgcaaatcagctatcaacaaagacaaagtaaatgtagcctaataaaaagc 4919  
Db 4861 ctcaaatgaattgcaaatcagctatcaacaaagacaaagtaaatgtagcctaataaaaagc 4920  
Qy 4920 aaggcatcaatgaagacaaccccttgttaaaggacttgaaaaagccgcttctgataaca 4979  
Db 4921 aaggcatcaatgaagacaaccccttgttaaaggacttgaaaaagccgcttctgataaca 4980  
Qy 4980 aaacccaaaaacgcgcgagtaactgtgggtgatttaaatgccgttgcccaaacacgcgtga 5039  
Db 4981 aaacccaaaaacgcgcgagtaactgtgggtgatttaaatgccgttgcccaaacacgcgtga 5040  
Qy 5040 cctttgcaagggtatcacaggcaacgcgttaaaaaactggcgagactttgaccatcaag 5099  
Db 5041 cctttgcaagggtatcacaggcaacgcgttaaaaaactggcgagactttgaccatcaag 5100  
Qy 5100 gtgggcaaacagacaccaataagctaaccgataataacatcggtgtgtagcaggtactg 5159  
Db 5101 gtgggcaaacagacaccaataagctaaccgataataacatcggtgtgtagcaggtactg 5160  
Qy 5160 atggtctcaactgtcaaaacttgcgaagacctaaacaaactcttaacagcgttaatgcaggtg 5219  
Db 5161 atggtctcaactgtcaaaacttgcgaagacctaaacaaactcttaacagcgttaatgcaggtg 5220  
Qy 5220 gcacccaaatgtagacaaaagcgtgtctttttagactcaagcgttcaagccaaaagcaa 5279  
Db 5221 gcacccaaatgtagacaaaagcgtgtctttttagactcaagcgttcaagccaaaagcaa 5280  
Qy 5280 acacccctgtctgaagtgcgaatgggctggacctgggtggcgaaggtcatcagtaattggtg 5339  
Db 5281 acacccctgtctgaagtgcgaatgggctggacctgggtggcgaaggtcatcagtaattggtg 5340  
Qy 5340 gcaaggcacaanaagataccgacgctgccaatgtacaacagtttaaacgaagtacgcaact 5399  
Db 5341 gcaaggcacaanaagataccgacgctgccaatgtacaacagtttaaacgaagtacgcaact 5400  
Qy 5400 tgttgggtcttgggttaattgataacgctgacggccaatcaggtaaacattgccc 5459  
Db 5401 tgttgggtcttgggttaattgataacgctgacggccaatcaggtaaacattgccc 5460  
Qy 5460 acatcaaaaaagaccccaattcaggttcatcatctaaccgcactgtcatcaaaagcaggca 5519  
Db 5461 acatcaaaaaagaccccaattcaggttcatcatctaaccgcactgtcatcaaaagcaggca 5520  
Qy 5520 cggtaacttggcgttaagggttaataacgataccgaaaaaacttgcacactgttggtatacaag 5579  
Db 5521 cggtaacttggcgttaagggttaataacgataccgaaaaaacttgcacactgttggtatacaag 5580  
Qy 5580 tggcggtggataaagacgcgcacgttaacgcgatttaagccaattgttgggtcaaaaccc 5639  
Db 5581 tggcggtggataaagacgcgcacgttaacgcgatttaagccaattgttgggtcaaaaccc 5640

Qy 5640 aaaaagatggcagcaaaaaagccctgtcgcgcacctataaacgccgaggttcagaccaact 5699  
Db 5641 aaaaagatggcagcaaaaaagccctgtcgcgcacctataaacgccgaggttcagaccaact 5700  
Qy 5700 atttgacaaacaaaccccgagaaagccattgacagaataaatagaacaaaggtatccgctctct 5759  
Db 5701 atttgacaaacaaaccccgagaaagccattgacagaataaatagaacaaaggtatccgctctct 5760  
Qy 5760 tccatgtcaacgatggcaatcaagagcctgtggtaacaagggcgtaacggcatttgactcaa 5819  
Db 5761 tccatgtcaacgatggcaatcaagagcctgtggtaacaagggcgtaacggcatttgactcaa 5820  
Qy 5820 gtgcctcaggcaagcactcagtgggcga taggtttccaggccaaagcagatgggtgaagccg 5879  
Db 5821 gtgcctcaggcaagcactcagtgggcga taggtttccaggccaaagcagatgggtgaagccg 5880  
Qy 5880 ccgttgcataaggcagacaaccccaagcaggaacaaacccatccatcccatcggtgataagc 5939  
Db 5881 ccgttgcataaggcagacaaccccaagcaggaacaaacccatccatcccatcggtgataagc 5940  
Qy 5940 cacaagccacagggcgatcaatccatccatcgccatcggttacaggaacaatgtgtagcagtaagc 5999  
Db 5941 cacaagccacagggcgatcaatccatccatcgccatcggttacaggaacaatgtgtagcagtaagc 6000  
Qy 6000 actctggtgccatcggcgacccaagcactgttaaggctgataaacggttacagttgggtga 6059  
Db 6001 actctggtgccatcggcgacccaagcactgttaaggctgataaacggttacagttgggtga 6060  
Qy 6060 ataacaacccagttaccgatgccactcaaacogatactgtttggtgggcaataacatca 6119  
Db 6061 ataacaacccagttaccgatgccactcaaacogatactgtttggtgggcaataacatca 6120  
Qy 6120 ccgtgacgcgaagtaactcggttgccttaggttcaaaactctgccatcagtcagtcagcacac 6179  
Db 6121 ccgtgacgcgaagtaactcggttgccttaggttcaaaactctgccatcagtcagtcagcacac 6180  
Qy 6180 acgcaggcacacaagccaaaaactctgacgcacagcaggttacaccaccacacagcaggtg 6239  
Db 6181 acgcaggcacacaagccaaaaactctgacgcacagcaggttacaccaccacacagcaggtg 6240  
Qy 6240 caacccgttacaggttaaaagccttgcgtgacaaaacggcgttgggtgcgtctccgtgggtg 6299  
Db 6241 caacccgttacaggttaaaagccttgcgtgacaaaacggcgttgggtgcgtctccgtgggtg 6300  
Qy 6300 cctcaggtgctgaacgcgctatccaaaatgtggcagcaggtgagtcagtcagtcacacaga 6359  
Db 6301 cctcaggtgctgaacgcgctatccaaaatgtggcagcaggttgggtgcagtcagtcacacaga 6360  
Qy 6360 ccgatgcggttcaatgggttagcaggttgcataaagccaccccaagcattgccaaacgaacca 6419  
Db 6361 ccgatgcggttcaatgggttagcaggttgcataaagccaccccaagcattgccaaacgaacca 6420  
Qy 6420 atgagcttgacacatcgtatccaccccaaaacgaataaaggccaaatgcagggatttcatcag 6479  
Db 6421 atgagcttgacacatcgtatccaccccaaaacgaataaaggccaaatgcagggatttcatcag 6480  
Qy 6480 cgatggcgatggcgctccatgcacaagcctacatcctggcagatcccatggttacccggg 6539  
Db 6481 cgatggcgatggcgctccatgcacaagcctacatcctggcagatcccatggttacccggg 6540  
Qy 6540 gtattggcaccaccaacacgcgtcaaggtgcggtggcgactgggactgtcgaaagcgtgtcggtga 6599  
Db 6541 gtattggcaccaccaacacgcgtcaaggtgcggtggcgactgtcgaaagcgtgtcggtga 6600  
Qy 6600 atggttcaatgggttatttaaaatcaatgggttcagccgatacccaagggccatgttagggcg 6659  
Db 6601 atggttcaatgggttatttaaaatcaatgggttcagccgatacccaagggccatgttagggcg 6660  
Qy 6660 cagttggtgcaggttttcaacttttaagccataaatacgaagattttacttaaaatacaat 6719  
Db 6661 cagttggtgcaggttttcaacttttaagccataaatacgaagattttacttaaaatacaat 6720



Qy	6720	ctcaccatagtggtataaaacagcatcagcatcagtcattactgctgattgtttt	6779
Db	6721	ctcaccatagtggtataaaacagcatcagcatcagtcattactgctgattgtttt	6780
Qy	6780	tatcaacttaaacacattttaccgcgtcaagtgatctctttccaccatgaccaaaatcgccatt	6839
Db	6781	tatcaacttaaacacattttaccgcgtcaagtgatctctttccaccatgaccaaaatcgccatt	6840
Qy	6840	gatcataggttaaacattatgagtaaaattttatcaatgtagtttttagatatggttaaaat	6899
Db	6841	gatcataggttaaacattatgagtaaaattttatcaatgtagtttttagatatggttaaaat	6900
Qy	6900	tgtgcattgacccaaaaatgacgcgattttatccgaaaaattcttgattatgacccgttga	6959
Db	6901	tgtgcattgacccaaaaatgacgcgattttatccgaaaaattcttgattatgacccgttga	6960
Qy	6960	cctgcaggtcgac	6972
Db	6961	cctgcaggtcgac	6973

RESULT 9

US-09-361-619-1

Sequence 1, Application US/09361619

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

APPLICANT: Sasaki, Ken

APPLICANT: Yang, Yan Ping

APPLICANT: Klein, Michel H.

TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE

FILE REFERENCE: 1038-921MIS:jb

CURRENT APPLICATION NUMBER: US/09/361.619

CURRENT FILING DATE: 1999-07-27

NUMBER OF SEQ ID NOS: 32

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 6973

TYPE: DNA

ORGANISM: Moraxella catarrhalis

US-09-361-619-1

Query Match	99.8%	Score 6961;	DB 17;	Length 6973;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 6972;	Conservative	0;	Mismatches	0;
			Indels	1;
			Gaps	1;

Qy	1	coatggaatggcgaggtgtcgtcgccgtatgatggcgatgacacccatttgccc	60
Db	1	coatggaatggcgaggtgtcgtcgccgtatgatggcgatgacacccatttgccc	60
Qy	61	catatctgacgatttgacatgtagtatttaacatgtgacatgatttaacattgttt	120
Db	61	catatctgacgatttgacatgtagtatttaacatgtgacatgatttaacattgttt	120
Qy	121	aatactgttgcattaccattaccataattttgagtaacgcatttagtaacgcatttggtaaaat	180
Db	121	aatactgttgcattaccattaccataattttgagtaacgcatttagtaacgcatttggtaaaat	180
Qy	181	cattgcgccctttatgttatcatatgaatagaattattatgatctctgattcatgt	240
Db	181	cattgcgccctttatgttatcatatgaatagaattattatgatctctgattcatgt	240
Qy	241	atcagaatggtgacttatgatgatgctacagtgattggttggttaatcacctctatg	300
Db	241	atcagaatggtgacttatgatgatgctacagtgattggttggttaatcacctctatg	300
Qy	301	attgatataattttgaaactaacttatgaacttaacccatgcttggttaattagca	360
Db	301	attgatataattttgaaactaacttatgaacttaacccatgcttggttaattagca	360
Qy	361	taatggtgaggtttttgaaaaatcacatcgcaatttcttactgttactacattgt	420





||||| aagtcggtgctgagtgctggaagttgcgaaggttaataataatggttgtagtgctg 3720 Db  
3720 gcaattgatggcacaaactcgcatcaccagagatgaaattggtttactggagactaaatggct 3779 Qy  
3721 gcaattgatggcacaaactcgcatcaccagagatgaaattggtttactggagactaaatggct 3780 Db  
3780 cacttgataaaagcacaaccccaacttaagcaagacggttaacgaggtggttaaaaga 3839 Qy  
3781 cacttgataaaagcacaaccccaacttaagcaagacggttaacgaggtggttaaaaga 3840 Db  
3840 ttaccaacttcaatcaggttgagattgccccaaaaacgcatgatgctgacagggcgga 3899 Qy  
3841 ttaccaacttcaatcaggttgagattgccccaaaaacgcatgatgctgacagggcgga 3900 Db  
3900 agatttatgatttaaaacgcaacttgaaaaaaacacagcagctgcacaaacagcac 3959 Qy  
3901 agatttatgatttaaaacgcaacttgaaaaaaacacgcaactgcacaaacagcac 3960 Db  
3960 aaactcattacacgaattctcagtagcagatgaacaaggtataaactttaagggttagta 4019 Qy  
3961 aaactcattacacgaattctcagtagcagatgaacaaggtataaactttaagggttagta 4020 Db  
4020 acccttactccagttatgacacctcaaaagacacctctgatgtcatcacctttgcaagtgaaa 4079 Qy  
4021 acccttactccagttatgacacctcaaaagacacctctgatgtcatcacctttgcaagtgaaa 4080 Db  
4080 acggcattacaccaaaggtataaagggtggtgctgctggtggcattgaccaaaccaaaag 4139 Qy  
4081 acggcattacaccaaaggtataaagggtggtgctgctggtggcattgaccaaaccaaaag 4140 Db  
4140 gcttaaccacgcctaagctgaccgtgggttaataaatggcaaggaattgtcatgaca 4199 Qy  
4141 gcttaaccacgcctaagctgaccgtgggttaataaatggcaaggaattgtcatgaca 4200 Db  
4200 gcaaaaatgctaaaaataccatcacaggaactaagcaacacctagctaatgttaccaatg 4259 Qy  
4201 gcaaaaatgctaaaaataccatcacaggaactaagcaacacctagctaatgttaccaatg 4260 Db  
4260 ataaagtagcgtacgacacacagacagggcgaataatacaaaagacgaagacaaaacc 4319 Qy  
4261 ataaagtagcgtacgacacacagacagggcgaataatacaaaagacgaagacaaaacc 4320 Db  
4320 gtgcgcgcagcattgtgtgctgtaagcagcaggttaacttgcagggcaatgggtgaag 4379 Qy  
4321 gtgcgcgcagcattgtgtgctgtaagcagcaggttaacttgcagggcaatgggtgaag 4380 Db  
4380 cgggtgacttctccacttatgacacgttgcacttgcggtggaatggcaccacgg 4439 Qy  
4381 cgggtgacttctccacttatgacacgttgcacttgcggtggaatggcaccacgg 4440 Db  
4440 ctaaggtgacctatgtagcacagcaaaaccagtaaagtgggtctatgtagtcaatgtgg 4499 Qy  
4441 ctaaggtgacctatgtagcacagcaaaaccagtaaagtgggtctatgtagtcaatgtgg 4500 Db  
4500 atgatacaaccattgaagttaaagataaaaaacttgcgttaaaaaaccacacattgacca 4559 Qy  
4501 atgatacaaccattgaagttaaagataaaaaacttgcgttaaaaaaccacacattgacca 4560 Db  
4560 gtactggcaaggtgcttaataaatttgcctaaagcaatcaagctactggcggtgcttg 4619 Qy  
4561 gtactggcaaggtgcttaataaatttgcctaaagcaatcaagctactggcggtgcttg 4620 Db  
4620 tcaaggccagtgatcgttgcctatctataacaccttatcttgcgcacatccaaactgcca 4679 Qy  
4621 tcaaggccagtgatcgttgcctatctataacaccttatcttgcgcacatccaaactgcca 4680 Db  
4680 aaggggcaagcagcaacactcagcaggtgatggatgctgtagtggaataaaggtca 4739 Qy  
4681 aaggggcaagcagcaacactcagcaggtgatggatgctgtagtggaataaaggtca 4740 Db  
4740 tctatgacgttaccgataaacaagtaactatcaagcccaaaaatgatggcacaggtgataaa 4799 Qy  
|||||

4741 tctatgacgttaccgataaacaagtaactatcaagccaaaaatgatggcacagttgataaaa 4800 Db  
4800 ccaaaagagttgcccacaaagcaaaactggctgcaccaagccaaacccagatggcacattgg 4859 Qy  
4801 ccaaaagagttgcccacaaagcaaaactggctgcaccaagccaaacccagatggcacattgg 4860 Db  
4860 ctcaaatgaattcaaatcagtcatttaacaaagaacaagtaaatgatgccaataaaaagc 4919 Qy  
4861 ctcaaatgaattcaaatcagtcatttaacaaagaacaagtaaatgatgccaataaaaagc 4920 Db  
4920 aaggcatcaatgaagacaaacgcttctgttaaaggacttgaaaaagcgcgttctgtataaca 4979 Qy  
4921 aaggcatcaatgaagacaaacgcttctgttaaaggacttgaaaaagcgcgttctgtataaca 4980 Db  
4980 aaacaaaaaacgccgcagtaactgtgttgatttaaatccgttgcctcccaaacacgcgtga 5039 Qy  
4981 aaacaaaaaacgccgcagtaactgtgttgatttaaatccgttgcctcccaaacacgcgtga 5040 Db  
5040 cctttgcaggggatacagggcacaaaggttaaaaaactggcgagactttgacacatacaag 5099 Qy  
5041 cctttgcaggggatacagggcacaaaggttaaaaaactggcgagactttgacacatacaag 5100 Db  
5100 gtgggcaaacagacacaaataagctaaacgataataacatcggtgtgtgtagcaggtactg 5159 Qy  
5101 gtgggcaaacagacacaaataagctaaacgataataacatcggtgtgtgtagcaggtactg 5160 Db  
5160 atggtctcactgtcaaaactgcccagacactaaacaaatttaaacacgcttaattgcaggtg 5219 Qy  
5161 atggtctcactgtcaaaactgcccagacactaaacaaatttaaacacgcttaattgcaggtg 5220 Db  
5220 gaacaaaattgatgacaaagcggtctctttgttagactcaagcgttcaagccaaagcaaa 5279 Qy  
5221 gaacaaaattgatgacaaagcggtctctttgttagactcaagcgttcaagccaaagcaaa 5280 Db  
5280 acacctgtgctaaatgcccattgggctggacgtgggtggcaaggttcacagtaactgtgg 5339 Qy  
5281 acacctgtgctaaatgcccattgggctggacgtgggtggcaaggttcacagtaactgtgg 5340 Db  
5340 gcaaaagcacaagaataccagacgtgcgaatgtacaacagttaaacgaagatcgcacact 5399 Qy  
5341 gcaaaagcacaagaataccagacgtgcgaatgtacaacagttaaacgaagatcgcacact 5400 Db  
5400 tgttggtctctgttaattgctgtgtaataacgtctgacgcaatcaggttaaacattgccc 5459 Qy  
5401 tgttggtctctgttaattgctgtgtaataacgtctgacgcaatcaggttaaacattgccc 5460 Db  
5460 acatcaaaaaagccccaaattcaggttctcatctaaacgcactgtcatcaaaagcagcca 5519 Qy  
5461 acatcaaaaaagccccaaattcaggttctcatctaaacgcactgtcatcaaaagcagcca 5520 Db  
5520 cgttacttggcggttaaggttaataacgataccgaaaaacttgcacttgggtggtatacaag 5579 Qy  
5521 cgttacttggcggttaaggttaataacgataccgaaaaacttgcacttgggtggtatacaag 5580 Db  
5580 tgggctgggttaagacggaacgcgttaacgcgtattttaagcaaatgtttgggtcaaaacc 5639 Qy  
5581 tgggctgggttaagacggaacgcgttaacgcgtattttaagcaaatgtttgggtcaaaacc 5640 Db  
5640 aaaaaagtgagcaaaaaagccccctgtcgcacttataacgcgcaggttcagaccaact 5699 Qy  
5641 aaaaaagtgagcaaaaaagccccctgtcgcacttataacgcgcaggttcagaccaact 5700 Db  
5700 atttgaccaaaccaccccgacagaagccattgacagaaataatgaacaaggtatccgctct 5759 Qy  
5701 atttgaccaaaccaccccgacagaagccattgacagaaataatgaacaaggtatccgctct 5760 Db  
5760 tccatgtcaacagatggcaatcaagagcgtgtgttacaagggcgtaacgagcattgactcaa 5819 Qy  
5761 tccatgtcaacagatggcaatcaagagcgtgtgttacaagggcgtaacgagcattgactcaa 5820 Db  
5820 gtgcctcagcgaacacactcagtgagtaggtttccagggccaagcagatgggtgaagccg 5879 Qy  
5821 gtgcctcagcgaacacactcagtgagtaggtttccagggccaagcagatgggtgaagccg 5880 Db







Db 4681 AAGGGCAAGCCAGCAACACTCAGCAGGCTATGTGGATGCTGATGGCAATTAAGGTCA 4740  
 Qy 4740 tctatgacagtaccgataaacaagtactatcaagcccaaaaaatgatggcacagttgataaaa 4799  
 Db 4741 TCTATGACAGTACCGATACCAAGTACTATCAAGCCAAATAATGATGGCACAGTTCGATAAAA 4800  
 Qy 4800 ccaagaagtgtccaaagacaaaacttggctggcccaagcccaaaacccacagatggacattgg 4859  
 Db 4801 CCAAGAACTTGCCAAAGACAACTTGGTGGCCAAAGCCCAAGCCCAAGCCAGATGGCATTGG 4860  
 Qy 4860 ctcaaatgaatcgaatcaatcagctattcaaaagaaacagtaaatgatgccaataaaaagc 4919  
 Db 4861 CTCAAATGAATGTCAAATCGTTCATTAACAGAACACAGTAATGATGCCAATAAAAAGC 4920  
 Qy 4920 aaggcatcaatgaagacaaacgccttgtttaaaggacttgaagaagccgcttcttgataaca 4979  
 Db 4921 AAGGCATCAATGAAGCAACACGCTTGTGTTAAAGGACTTTGAAAAGCCGCTTCTGTGATAACA 4980  
 Qy 4980 aaacccaaaacccgcagtaactgtgggtgatttaaatgcggttgcacaaacacccgtga 5039  
 Db 4981 AAACCAAAACCGCGCAGTAACTGTGGGTGATTTAAATGCCGTTGCCCAAAACACCGCTGA 5040  
 Qy 5040 cctttgcagggtgatacagcacaacgcgtcaaaacactggcgagactttgacatacaag 5099  
 Db 5041 CTTTGCAGGGGATACAGGCACAAACGGCTAAAAAACCTGGCGGAGACTTTGACATCAAAAG 5100  
 Qy 5100 gtgggcaaacagacacataagtaaacgataataaacatcggtgtgtagcagtgactg 5159  
 Db 5101 GTGGCAAAACAGACACAAATAAGCTAACCGATAATAACATCGGTGTGGTAGCAGTACTG 5160  
 Qy 5160 atggcttcactgtcaaaactgtccaaagacctaaccacttcttaacagcgtttaatgcagggtg 5219  
 Db 5161 ATGGCTTCACTGTCAAACTTGGCCAAAGACCTAACCAATCTTAACACGCTTAATGACAGGTG 5220  
 Qy 5220 gacccaaaattgatgacaaaggcgtgtctttgttagactcaagcgggtcaagccaaagcaa 5279  
 Db 5221 GCACCAAAATGTATGACAAAGGCGTGTCTTTGTAGACTCAACGGGTCAAGCCCAAGACAA 5280  
 Qy 5280 acacccctgtactgaatgccaatggcgtgacgtggtggcaagggtcaatgaatgtgg 5339  
 Db 5281 ACACCCCTGTGCTAAGTGCCAAATGGGCTGGACCTGGGTGGCAAGGTCAATGATGG 5340  
 Qy 5340 gcaaggcacaacaaatgacoga-cgctgccaatgtacaaacagttaaacgaaglacgcaac 5398  
 Db 5341 GCAAGGACACAAAAGATACCGACCGCTGCCAATGTACAAACAGTTAAACGAAGTACGCAAC 5400  
 Qy 5399 ttgttggtctgtgaatgctgtgtaataagctgaacggtacggttaaacattggc 5458  
 Db 5401 TTGTTGGGTCTTGGTAATGCTGTTAATGTAATGTAACGCTGACGGCAATCAGGTAAACATTGCC 5460  
 Qy 5459 gacatcaaaaagacccaaatccaggttcatcatcaacccgactgtcatcaagaagcgc 5518  
 Db 5461 GACATCAAAAAGACCCAAATTCAGTTTCAATCATCAACCGCACTGTCAACAAAGCAGGC 5520  
 Qy 5519 acggtacttgcgttaaggtaataagataaccgaaacacttgcactgtggtgatacaa 5578  
 Db 5521 ACGGTACTTGGCGTAAAGGTAATAACGATACCCGAAACACTTGGCACTGTGTTATACAA 5580  
 Qy 5579 gtggcgctggaataaagacgcaacgctaacggtatttaagcaatgtttgggtcaaaacc 5638  
 Db 5581 GTGGCGGTGGATTAAGACGGCAACGCTAACGGCGATTAAAGCAATGTTGGGTCAAAAACC 5640  
 Qy 5639 caaaaagatggcagcaaaaagccctgtcgcacttataacgcgcaggtcagaccacac 5698  
 Db 5641 CAAAAGATGGCAGCAAAAAGCCCTGTGCCCACTTATAACGGCGCAGGTGACACCAAC 5700  
 Qy 5699 tattgaccaacaacccgcagaagccattgacagaaataatgaacaaaggtatccgcttc 5758  
 Db 5701 TATTTGACCAACAACCCGCCAAGCCATTCACAGAAATAATGAACAAGGTATCCGCTTC 5760  
 Qy 5759 ttccatgtcaacgattggaatcaagagcgtgtgtacaaggcgtaacggcattgactca 5818  
 Db 5761 TTCCATGTCAAGGATGGCAANTCAAGAGCCTGTGGTACAAAGGCGTAAACGGCATGACTCA 5820

Qy 5819 atgacctcaggcaagcactcagtggcqataggtttccaggccaagcagatggtgaagcc 5878  
 Db 5821 AGTGCTCAGCGCAAGCACTCAGTGGCGATAGGTTTCCAGGCCAAGCAGATGTTGAAGCC 5880  
 Qy 5879 gccgttgccatagggcagacaaacccaagcaggcaacccaatccatccatcggtgataaac 5938  
 Db 5881 GCCGTTGCCATAGGCAGACAAACCAAGCAGGCAACCAATCCCATCGCATCGGTGATAAC 5940  
 Qy 5939 gcacaagccagggcgatcaatccatcgccatcggttacaggcaatggttagcaggtgaag 5998  
 Db 5941 GCACAAAGCCAGGGCGATCAATTCATCGCCATCGGTACAGGCAATGTTGTTAGCAGTAAG 6000  
 Qy 5999 cactctgtgcatcgccgacccaagcactgtttaaggctgataaacaggttacagtgtggtt 6058  
 Db 6001 CACTCTGTTGCCATCGCGACCCCAAGCACTGTTAAAGCTGATAACAGTTACAGTGTGGGT 6060  
 Qy 6059 aataacaacagctttacccgatgcacccaacccgatgtcttgggtgggcaataaacatc 6118  
 Db 6061 AATAACAACCAAGTTTACCGATGCCACTCAAAACCGATGCTTTGGTGTGGCGCAATAACATC 6120  
 Qy 6119 accgtgacggaagttaactcggttgccttaggtttcaaaactctgccaatcagtcaggcaca 6178  
 Db 6121 ACCGTGACCGAAAGTAATCTCGTTGCCCTTAGGTTCAAACTCTGCCATCAGTGCAGGCACA 6180  
 Qy 6179 cacgcaggcacacaagccaaaatactgacgcagcagcaggtacacacacacacagcaggt 6238  
 Db 6181 CACGCAGCACACAAAGCCAAAATACTGACGGCACAGCGGTACAAACCCACACAGCAGGT 6240  
 Qy 6239 gcaaccggtacggttaaaagccttgcctggacaaacgcggttgggtgctcctcggtgggt 6298  
 Db 6241 GCAACCGGTAGCGTTAAAGGCTTGTGACAAACCGCGGTGGTGGCGGTCTCGGTGGGT 6300  
 Qy 6299 gctcaggtgtgaacgcgcgtatccaaaatgtgcagcaggtgaggtcagtcagtcacacagc 6358  
 Db 6301 GCCTCAGGTGTGAACGCCGTATCCAAAATGTGCAGCAGGTGAGGTGAGTGCACCCACAGC 6360  
 Qy 6359 accgatcggtcaatggtagcagttgtacaaagccaccccaaaagcattgccaagcgaacc 6418  
 Db 6361 ACCGATGCGGTCAATGTTAGCAGCTTGTACAAAGCCACCCAAAGCATTTGCCAAGCAACC 6420  
 Qy 6419 aatgagcttgacatcgatccacccaaacaaataaggccaaatgcagggtatttcataca 6478  
 Db 6421 AATGAGTTGACATCGTATCCACCAAAACGAAATAAGGCCAATGCAAGGGATTTTCATCA 6480  
 Qy 6479 gcatgagcagtcgctccatgcacaaagcctacattcctgagcagatccatggttaccggg 6538  
 Db 6481 GCGATGGCGATGGCGTCCATGCGCACAAAGCCTACATTCCTGGCAGATCCATGTTACCAGG 6540  
 Qy 6539 ggtattgcccacccacaacgggtcaaggtgcgggtggcagtgggactgtcgaagcgtcggtat 6598  
 Db 6541 GGTATTGCCCACCAACACGCTCAAGGTGGGTGGCAGTGGGACTGTCCAAAGCTGTCGGAT 6600  
 Qy 6599 aatggtcaatgggtatttaaaatcaatggttcagccgatcccaagggccatgtaggggcg 6658  
 Db 6601 AATGSGTCAATGGGTATTTAAATCAATGTTTTCAGCCGATATCCCAAGGCCATGTAGGGGCG 6660  
 Qy 6659 gcaagttggtgaggttttcacttttaagccataaaatcgcaagatttttacttaaaatcaa 6718  
 Db 6661 GCAGTTGGTGCAGGTTTTCATTTTAAGCCATAAATGCAAGATTTTACTTAAATAATCAA 6720  
 Qy 6719 tctcaccatagttgtat-aaaacagcatcagcatcagtcataattactgactgactggttt 6777  
 Db 6721 TCTCACCATAAGTTGATAAAGCAGCATCAGCATAGTCATATTTACTGATGCTGATGTTT 6780  
 Qy 6778 ttatcacttaaacactttaccgctcaagtgattctcttccaccatgaccacaaatcgcca 6837  
 Db 6781 TTTATCACTTTAAACCAATTTTACCGCTCAAGTGATTTCTTTTCCACCATGACCAANTCGCCA 6840  
 Qy 6838 ttgatcataggttaaaccttattgagttaaattttatcaatgtagttgtagtatggttaaa 6897  
 Db 6841 TTGATCATAGTAAACTTATTGATTAATTTTATCAATGATGTTGTTAGATATGGTTAA 6900



QY 6898 attgtgcatgaccacaaatgaccgattttatcccgaaaaatttctgattatgacccgtt 6957  
|||||  
Db 6901 ATTGTGCCATTGACCAAAATGACCGATTATCCCGAAAATTTCTGATTATGATCCGTT 6960  
QY 6958 gacctgcaggtcgac 6972  
|||||  
Db 6961 GACCTGCGAGTGCAC 6975  
  
RESULT 11  
US-08-431-718A-1  
; Sequence 1, Application US/08431718A  
; GENERAL INFORMATION:  
; APPLICANT: Sasaki, Ken  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER  
; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Slim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/431.718A  
; FILING DATE: 01-MAY-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-429  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6975 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-431-718A-1  
  
Query Match 99.5%; Score 6939; DB 8; Length 6975;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6972; Conservative 0; Mismatches 0; Indels 3; Gaps 3;  
  
QY 1 ccatgatatggcgagggtgtgtcgtccgtccgtatgtagggcgatgacacccattggccc 60  
Db 1 CCATGGATATGGCAGGTGTGTCGCGCTGCCGTATGATGGCGATGACACCCATTGGCCC 60  
  
QY 61 catatctgtacagatttgacatgtgatattgatttaacatgtgacatgatttaacattgttt 120  
Db 61 CATATCTGTACGATTGTGACATGTGATGATTAAACATGTGACATGATTAAACATTGTTT 120  
  
QY 121 aatactgttgccatcattaccataatttagtaacgcatttagtaacgcaatttgtaaaaaa 180  
Db 121 AATACTGTGGCCATCATTAACCAATATTAGTAACGCATTATGACGCATTGTGAAAAAT 180  
  
QY 181 cattgcccctttatgtgtatcatatgaatagaatattatgattgtatctgattattgt 240  
Db 181 CATTCGCCCTTTATGTCATCAATGAATAGATATATGATTGATGATCTGATTATTCT 240  
  
QY 241 atcagaatgggtgatgctatatgatgatgcctacgaggttgatttgggttaatacactctatg 300  
|||||

Db 241 ATCAGAAATGGTGATGCTATATATGATGATGCTACGAGTTGATTTGGGTTAATCACCTCTATG 300  
QY 301 atttataatttttgaactaatctattgacttaaatcaccataatggttataattagca 360  
|||||  
Db 301 ATTTCATATATTTTGAACAACTAATCTATTTGACTTAAATCACCATATGGTTTATATTTAGCA 360  
QY 361 taatgtaggctttttgttaaaaaatcacatcgcaatattgttctactgttactaccatgt 420  
|||||  
Db 361 TAATGGTAGGCTTTTGTAAAAATCACATCGCAATATTCTTCTACTGTTACTACCATGCT 420  
QY 421 tgaatgacgatcccaatcacagattcattcaagtgtatgtgtttatatacgcacattta 480  
|||||  
Db 421 TGAATGACGATCCCAATCACCAGATTCAATCAAGTGTATGTTTGTATACGCCACCATTTA 480  
QY 481 cccatattttcaaatcaaatgcctatgtcagcatgtatcatttttttaaggtaaacac 540  
|||||  
Db 481 CCTAATTTTCAATCAAAATGCCATGTCACGATGATCATTTTAAAGGTAAACAC 540  
QY 541 catgaatcacatctataaagtcatcttttaacaaagccacagcacatttatggcagtggc 600  
|||||  
Db 541 CATGAATCACATCTATAAAGTCTATCTTTAACAAAGCCACAGGCACATTTATGCGAGTGC 600  
QY 601 agagtacgccaatccacagcac-gggggggggtagctgtctacagggcgaagtggca 659  
|||||  
Db 601 AGAGTACGCCAAATCCACAGCGGGGGGGGTAGCTGTGCTACAGGGCAAGTTGGCA 660  
QY 660 gtgtatgcaactctgagctttgcccgtattgcgcgcgtcgtcgtcctcgtgacgtgca 719  
|||||  
Db 661 GTGTATGCACTCTGAGCTTTGCCCGTATTGGCCGCGCTCGCTCTCTCTGTCGTCGTCG 720  
QY 720 cgtcagtgaggcagtgcttatgctcaaaaaaagataaccaaacatctcgaattgggaa 779  
|||||  
Db 721 CGCTCAGTGGCAGTCTTATGCTCAAAAAAAGATACCAACATATCGCAATTTGGTGAAC 780  
QY 780 aaaaacagcgaagcagctcagcactgcaagcgagcgtgtatcgagcattgtctatg 839  
|||||  
Db 781 AAAACGAGCCAAAGCGCTCAGGCATGCCAAGCGGACGGTGTATGAGGCATTTGCTATTG 840  
QY 840 gtgaaaatgtcaacgcacagcggttcaagccatcgccatcgtagtagtaataaaactg 899  
|||||  
Db 841 GTGAAAATGCTAACGCACAGCGCGGTCAAGCCATCGCCATCGGTAGTAGTAATAAATCTG 900  
QY 900 tcaatggaagcagtttggataagataggtaccgatgctacgggtcaagagtcacatcgcca 959  
|||||  
Db 901 TCAATGGAAGCAGTTTGGATAAGATAGTAGTACCGGTCTACGGGTCAAGAGTCCATCGCCA 960  
QY 960 tcggtgggtatgtaaaaggctagtggtgagcgtcgtcgtatcccatcggtatgtagcttac 1019  
|||||  
Db 961 TCGGTGGTGTATGTAAGGCTAGTGTGATGTCCTCGATTGCCATCGGTAGTGTGATGCTTAC 1020  
QY 1020 atttctgtatcagcatggttaactcctaaacatccgaaggtactctgtatgaacgactta 1079  
|||||  
Db 1021 ATTTCTGTATCAGCATGTTAATCTTAACATCCGNAAGGTACTCTGATTACGATCTTTA 1080  
QY 1080 ttaacggccatgcagttattaaagaataatcgaagctcaagagtaaatgaltgaaataata 1139  
|||||  
Db 1081 TTAACGGCCATGCAGTATTAAAAAGAAATACGAAGCTCAAAAGGATATATGATGTAATAATA 1140  
QY 1140 gacgcacaaacgcaagcgagacagccagtaactgtagtggagggccatgcatatgacagg 1199  
|||||  
Db 1141 GACGCACAAACCCCAAGCGGACACGCGAGTACTGCAAGTGGGAGCCATGTCATATGCAACAG 1200  
QY 1200 gtcaatttttccaaacgctttgtgtaacgggcaacagctaaaagtgctcttctctggcag 1259  
|||||  
Db 1201 GTCAATTTTCCAAACGCCCTTTGGTACACGGGCAACAGCTAAAAGTGCCTATTCTCTGGCAG 1260  
QY 1260 tgggttctgcccacagccgaggggccaatctacaatcgctattgttctgtatgcaacat 1319  
|||||  
Db 1261 TGGGTCTTCCGCCACACAGCCGAGGGCCAACTTACAATCGCTATTGGTTCTGTATGCAACAT 1320  
QY 1320 ctagctcgttgggagcgatagcccttggtaggggtactcgtcgtcagctacagggcagta 1379  
|||||  
Db 1321 CTAGTCGTGGGAGCGATAGCCCTTGGTGCAGGTACTCGTCTCAGCTACAGGGCAGTA 1380







QY	5759	ttccatgtcaacgcatggcgaatcaagagcctgtgtgtacaaggcgtaacgagcattgactca	5818
DB	5761	ttctcattgtcaacgattggcgaatcaagagcctgtgtgtacaaggcgtaacgagcattgactca	5820
QY	5819	agtgctcagggaagcaactcagtggtggttaggtttccaggccaagcgagatggtgaagcc	5878
DB	5821	agtgctcctcagggaagcactcagtggtggttaggtttccaggccaagcgagatggtgaagcc	5880
QY	5879	gccgtttgcattagcgagacaacccaagcagcgcaaccaatccatcgccatcggtgataac	5938
DB	5881	gccgtttgcattagcgagacaacccaagcagcgcaaccaatccatcgccatcggtgataac	5940
QY	5939	gcacaagccaagggcgatcaatccatcgccatcggtcagggcaaatgtggtagcaggaag	5998
DB	5941	gcacaagccaagggcgatcaatccatcgccatcggtcagggcaaatgtggtagcaggaag	6000
QY	5999	cactctggtgcattcggcgagcccaagcactgttaagctgtataacagttcacagttcggt	6058
DB	6001	cactctggtgcattcggcgagcccaagcactgttaagctgtataacagttcacagttcggt	6060
QY	6059	aatacaaacccagttaccgatgccactcaaacccgatctcttggtgtggggcaataacatc	6118
DB	6061	aatacaaacccagttaccgatgccactcaaacccgatctcttggtgtggggcaataacatc	6120
QY	6119	acggtgacggaagtaactcgtgtcccttaggttcaaacctctgcactcagtcagcgaca	6178
DB	6121	acggtgacggaagtaactcgtgtcccttaggttcaaacctctgcactcagtcagcgaca	6180
QY	6179	cacgagggcacacaagcccaaaaaactcagcgccagcaggtacaaaccaaccaagcaggt	6238
DB	6181	cacgagggcacacaagcccaaaaaactcagcgccagcaggtacaaaccaaccaagcaggt	6240
QY	6239	gcaacggtagcgttaagagctttcgtgacaaaacggcggttggtcggtctccgtgggt	6298
DB	6241	gcaacggtagcgttaagagctttcgtgacaaaacggcggttggtcggtctccgtgggt	6300
QY	6299	gcctcaggtgctgaacgcgctatcccaaaatgtggcgagcaggtgaggtcagtcgcaaccagc	6358
DB	6301	gcctcaggtgctgaacgcgctatcccaaaatgtggcgagcaggtgaggtcagtcgcaaccagc	6360
QY	6359	acgatcgtcgaatgtagcaggtgtacaagccaccgaagcattgcgaacgcaacc	6418
DB	6361	acgatcgtcgaatgtagcaggtgtacaagccaccgaagcattgcgaacgcaacc	6420
QY	6419	aatgagcttgaccatcgtatccaccacaaaacgaaaaataaggccaatgcagggaattcatca	6478
DB	6421	aatgagcttgaccatcgtatccaccacaaaacgaaaaataaggccaatgcagggaattcatca	6480
QY	6479	gcgatggcgatggcgctccatgcgaagcctacatctcctgcagatccatggttacgggg	6538
DB	6481	gcgatggcgatggcgctccatgcgaagcctacatctcctgcagatccatggttacgggg	6540
QY	6539	ggattgccaccacacggtcaggtgcggtggcggtggagctgtcgaagctgcggtat	6598
DB	6541	ggattgccaccacacggtcaggtgcggtggcggtggagctgtcgaagctgcggtat	6600
QY	6599	aatggccaatgggtattttaaaalcaaatggttcagccgataccccaggccatgtagggcg	6658
DB	6601	aatggccaatgggtattttaaaalcaaatggttcagccgataccccaggccatgtagggcg	6660
QY	6659	gcagttggtgcaggttttccacttttaagccataaatcgcgaagatttacttaaaaaatcaa	6718
DB	6661	gcagttggtgcaggttttccacttttaagccataaatcgcgaagatttacttaaaaaatcaa	6720
QY	6719	ttccaccatagttgat-aaacaagcatcagcatcagtcattactgactgactgattt	6777
DB	6721	ttccaccatagttgat-aaacaagcatcagcatcagtcattactgactgactgattt	6780
QY	6778	tttatcacttaaacccatttaccgctcaagtgattctctcttccaccatgacccaatcgcca	6837
DB	6781	tttatcacttaaacccatttaccgctcaagtgattctctcttccaccatgacccaatcgcca	6840

QY	6938	ttgatcatagggtaaacctattgagtaaatatttatcaatgtaagttggttagatatacgttataaa	6899	
DB	6941	TTGATCATAGGTAACCTTATTGAGTAAATTTTATCAATGTTAGTTGTTAGATATGGTTAAA	6900	
QY	6998	atgtgaccattgacccaaaaagtcaggattatcccgaaaaatttctgattatgatccgtt	6957	
DB	6901	ATTGTGCCATTGACCAAAAANTGACCGATTATCCGAAATTTCTGATTATGATCCGTT	6960	
QY	6958	gacctgcaggctgcag 6972		
DB	6961	GACCTGCAGCTGCAC 6975		
RESULT 12				
US-08-431-718C-1				
; Sequence 1, Application US/08431718C				
; GENERAL INFORMATION:				
; APPLICANT: Ken, Sasaki				
; APPLICANT: Robin, Harkness E.				
; APPLICANT: Loosmore, Sheena M.				
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN O				
; TITLE OF INVENTION: MORAXELLA				
; FILE REFERENCE: 1038-429 MIS				
; CURRENT APPLICATION NUMBER: US/08/431,718C				
; CURRENT FILING DATE: 1995-05-01				
; NUMBER OF SEQ ID NOS: 4				
; SOFTWARE: Patent In Ver. 2.1				
; SEQ ID NO 1				
; LENGTH: 6975				
; TYPE: DNA				
; ORGANISM: Moraxella catarrhalis				
US-08-431-718C-1				
Query Match 99.5%; Score 6939; DB 8; Length 6975;				
Best Local Similarity 100.0%; Pred. No. 0;				
Matches 6972; Conservative 0; Mismatches 0; Indels 3; Gaps				
QY	1	ccatggatatggcagggtgtcgcctgcctgctgatgacgatgcgatgacacccatttgcgc 60		
DB	1	ccatggatatggcagggtgtcgcctgcctgctgatgacgatgcgatgacacccatttgcgc 60		
QY	61	catatctgtacgatttgacatgtgatatttaaacatgtgcacatgatttaacattgttt 120		
DB	61	catatctgtacgatttgacatgtgatatttaaacatgtgcacatgatttaacattgttt 120		
QY	121	aatactgttgcacatcatcacataatttagtaacgcatttagtaacgcatttggtaaaaaat 180		
DB	121	aatactgttgcacatcatcacataatttagtaacgcatttagtaacgcatttggtaaaaaat 180		
QY	181	catgtgcgcctttatgtgatcatatgaatagaatatattatgatgtatctgtatttgt 240		
DB	181	catgtgcgcctttatgtgatcatatgaatagaatatattatgatgtatctgtatttgt 240		
QY	241	atcgaatgggtgatgcgtatgatgatgcctacagttgatttgggttaatcacctctatg 300		
DB	241	atcgaatgggtgatgcgtatgatgatgcctacagttgatttgggttaatcacctctatg 300		
QY	301	atttgatataatttgaacaaactaatctattgtacttaaacatcattatgtttataatttagca 360		
DB	301	atttgatataatttgaacaaactaatctattgtacttaaacatcattatgtttataatttagca 360		
QY	361	taatggtaggctttttgaaaaatcacatgcgcaatatgttctcactgtttactaccattgct 420		
DB	361	taatggtaggctttttgaaaaatcacatgcgcaatatgttctcactgtttactaccattgct 420		
QY	421	tgaatgcagatcccaatcacacagatttcaatcaagtgtatgtttgtatagcacacattta 480		
DB	421	tgaatgcagatcccaatcacacagatttcaatcaagtgtatgtttgtatagcacacattta 480		
QY	481	cctaattatttcaatcaaatgcctatgtcagcagtgatcatcttttttaagggttaaacacc 540		

Db 481 cccataatttcaatacaaatgcctatgtcagcgatgtatcatttttttaagggttaaacac 540  
Qy 541 catgaatcacatctataaagtcattctttaacaaagccacagcgacatttatggcagtgcc 600  
Db 541 catgaatcacatctataaagtcattctttaacaaagccacagcgacatttatggcagtgcc 600  
Qy 601 agagtacgcaaaatccacacagcac-gggggggggtagctgtgctacagggcgaagtggca 659  
Db 601 agagtacgcaaaatccacacagcac-gggggggggtagctgtgctacagggcgaagtggca 660  
Qy 660 gtgtatgaactctgagctttgcccgtattgcgcgctgcgtctgctcgtgcatcggtcaca 719  
Db 661 gtgtatgaactctgagctttgcccgtattgcgcgctgcgtctgctcgtgcatcggtcaca 720  
Qy 720 cgcctagtggcagtgcttatgctcaaaaaaagataccaaacatatcgcaattggtgaac 779  
Db 721 cgcctagtggcagtgcttatgctcaaaaaaagataccaaacatatcgcaattggtgaac 780  
Qy 780 aaaaaccagccaagcgcctcagcgactccaagggcgagcggtgatcgagccaattgctattg 839  
Db 781 aaaaaccagccaagcgcctcagcgactccaagggcgagcggtgatcgagccaattgctattg 840  
Qy 840 gtgaaaatgctaaacgcacagggcggtcaagccatcgccatcggttagtagtaataaaaactg 899  
Db 841 gtgaaaatgctaaacgcacagggcggtcaagccatcgccatcggttagtagtaataaaaactg 900  
Qy 900 tcaatggaagcagtttggaataagtagtaccgatgctacgggtcaagagtcacatcgcca 959  
Db 901 tcaatggaagcagtttggaataagtagtaccgatgctacgggtcaagagtcacatcgcca 960  
Qy 960 tcggtggtgatgtaaaagctagtggtgatgcctcgattgcctcgattgcctcgattgccttac 1019  
Db 961 tcggtggtgatgtaaaagctagtggtgatgcctcgattgcctcgattgcctcgattgccttac 1020  
Qy 1020 attgctgtgacagctgttaactcctaaacatccgaaggtactctgattaaacgatctta 1079  
Db 1021 attgctgtgacagctgttaactcctaaacatccgaaggtactctgattaaacgatctta 1080  
Qy 1080 ttaacggccatgcagttataaagaataacgaagctcaaaagataaattgatgtaaaatata 1139  
Db 1081 ttaacggccatgcagttataaagaataacgaagctcaaaagataaattgatgtaaaatata 1140  
Qy 1140 gacgcacaacgcgaagcggaacacgcagctgagtgaggaggaagcattgcatatgcacagg 1199  
Db 1141 gacgcacaacgcgaagcggaacacgcagctgagtgaggaggaagcattgcatatgcacagg 1200  
Qy 1200 gtcatttttccaaacgctttgtgtacagggccaacagctaaaaagtgcctattccttggcag 1259  
Db 1201 gtcatttttccaaacgctttgtgtacagggccaacagctaaaaagtgcctattccttggcag 1260  
Qy 1260 tgggtcttgcgcgcacagccagcgaggccaatctacaatcgctattggttctgtagcaaacat 1319  
Db 1261 tgggtcttgcgcgcacagccagcgaggccaatctacaatcgctattggttctgtagcaaacat 1320  
Qy 1320 ctagctcgttgggagcgatagcccttgggtcaggtactcgtgctcagctacagggcgagta 1379  
Db 1321 ctagctcgttgggagcgatagcccttgggtcaggtactcgtgctcagctacagggcgagta 1380  
Qy 1380 ttgcctcaggttcaagggtctgttgcactcagagtgataataattctagaccggccctata 1439  
Db 1381 ttgcctcaggttcaagggtctgttgcactcagagtgataataattctagaccggccctata 1440  
Qy 1440 caccaaataccagggcactagaccaccaagtttcaagccaccaataataacgaagcggggtc 1499  
Db 1441 caccaaataccagggcactagaccaccaagtttcaagccaccaataataacgaagcggggtc 1500  
Qy 1500 cactttccattggttagtaactctataacagtaaaaaatcaataatgctcgggtcaggtgtta 1559  
Db 1501 cactttccattggttagtaactctataacagtaaaaaatcaataatgctcgggtcaggtgtta 1560  
Qy 1560 ataaaaccgagtcggttcaatgtgacagctagaaagcggtggtgaagtgggctaagagc 1619  
Db 1561 ataaaaccgagtcggttcaatgtgacagctagaaagcggtggtgaagtgggctaagagc 1620

Qy 1620 gtagaattacttttcaggggtgatgataacagctactgacgtataaaatagtttggataata 1679  
Db 1621 gtagaattacttttcaggggtgatgataacagctactgacgtataaaatagtttggataata 1680  
Qy 1680 ctttaactatttaaagggtggtgcagagacccaacgcatataaccgataataatcctggtg 1739  
Db 1681 ctttaactatttaaagggtggtgcagagacccaacgcatataaccgataataatcctggtg 1740  
Qy 1740 taaaagaggtgataaagtggtctgaaagttaaacttgctaaaactttaaacatctta 1799  
Db 1741 taaaagaggtgataaagtggtctgaaagttaaacttgctaaaactttaaacatctta 1800  
Qy 1800 ctgaggtgaaatacaactacattaaatgccacaacacagtttaaggtaggtagtagtagta 1859  
Db 1801 ctgaggtgaaatacaactacattaaatgccacaacacagtttaaggtaggtagtagtagta 1860  
Qy 1860 gtactacagctgaattatttgatgtagtatttaacctttaccagcccaatacagggcagtc 1919  
Db 1861 gtactacagctgaattatttgatgtagtatttaacctttaccagcccaatacagggcagtc 1920  
Qy 1920 aaagcacaagcaaaacgcgtctatggcgttaatgggtgaaagttaacttaataatgcagaaa 1979  
Db 1921 aaagcacaagcaaaacgcgtctatggcgttaatgggtgaaagttaacttaataatgcagaaa 1980  
Qy 1980 caacagcagcgaatcggcactactcgtattaccagagataaaaattgggttctcgagatg 2039  
Db 1981 caacagcagcgaatcggcactactcgtattaccagagataaaaattgggttctcgagatg 2040  
Qy 2040 gtgatgttgatgaaaaaacaagcacacatatattggataaaaaacaacttaagtgggtagtg 2099  
Db 2041 gtgatgttgatgaaaaaacaagcacacatatattggataaaaaacaacttaagtgggtagtg 2100  
Qy 2100 ttgcaattaccatagacaatggcattgagtgaggttaataaaaaaatcagtaactcttgcca 2159  
Db 2101 ttgcaattaccatagacaatggcattgagtgaggttaataaaaaaatcagtaactcttgcca 2160  
Qy 2160 aaggtagcagtgctaaacgagtcggttaccatcgcaacagctcaaaagccgccaagcctactt 2219  
Db 2161 aaggtagcagtgctaaacgagtcggttaccatcgcaacagctcaaaagccgccaagcctactt 2220  
Qy 2220 taaacgagggcgctggcactcagtgctcacactactgaaataatcagttgtagctaaagatg 2279  
Db 2221 taaacgagggcgctggcactcagtgctcacactactgaaataatcagttgtagctaaagatg 2280  
Qy 2280 gcaatgtttaccgccccaaacttacaacattggcggtgaaaaaccacgagcttaacagtgatg 2339  
Db 2281 gcaatgtttaccgccccaaacttacaacattggcggtgaaaaaccacgagcttaacagtgatg 2340  
Qy 2340 gcaatgtgataaaatttagtggttaagggtagtggtacgaaacaatagcttagttaccgccc 2399  
Db 2341 gcaatgtgataaaatttagtggttaagggtagtggtacgaaacaatagcttagttaccgccc 2400  
Qy 2400 aacatttggcgaagctataataaagaatcgaacggtgacaggtgctgctctacaaagct 2459  
Db 2401 aacatttggcgaagctataataaagaatcgaacggtgacaggtgctgctctacaaagct 2460  
Qy 2460 ttaccgtttaaagaagaagacgatgatgacgccaacgctcaccgtggtctaaaagatacga 2519  
Db 2461 ttaccgtttaaagaagaagacgatgatgacgccaacgctcaccgtggtctaaaagatacga 2520  
Qy 2520 caaaaaatgccggcgcgagtcagcatctttaaactcaaaaggttaaaacggtctaacgggtg 2579  
Db 2521 caaaaaatgccggcgcgagtcagcatctttaaactcaaaaggttaaaacggtctaacgggtg 2580  
Qy 2580 ctacaaaaaagaatgggtacgggttacccttgggttagccaagatgagcgtctgaccattg 2639  
Db 2581 ctacaaaaaagaatgggtacgggttacccttgggttagccaagatgagcgtctgaccattg 2640  
Qy 2640 gcaaaagcaccttaaaacaacgatggttgactgttaagatataccaacgaaacaaatcaag 2699  
Db 2641 gcaaaagcaccttaaaacaacgatggttgactgttaagatataccaacgaaacaaatcaag 2700

QY 2700 tcggtgctaatgacattataatttactaatgtgaattggttagtataccaggtactggtcattg 2759  
|||||  
Db 2701 tcggtgctaatgacattataatttactaatgtgaattggttagtataccaggtactggtcattg 2760  
|||||  
QY 2760 caaataccgctgcgattaccagagataaaattggctttgctgtgctgtgctgattggtgagttg 2819  
|||||  
Db 2761 caaataccgctgcgattaccagagataaaattggctttgctgtgctgtgctgattggtgagttg 2820  
|||||  
QY 2820 atacaacaaacaccttactctgtatcaagacaaagctacacagttggcaattgtaagattacca 2879  
|||||  
Db 2821 atacaacaaacaccttactctgtatcaagacaaagctacacagttggcaattgtaagattacca 2880  
|||||  
QY 2880 acactggcattacagcaggtgttaagccatcacagggctgtccccaacactgccttagca 2939  
|||||  
Db 2881 acactggcattacagcaggtgttaagccatcacagggctgtccccaacactgccttagca 2940  
|||||  
QY 2940 ttgccgatcaaatgtagccgaacatagaactgggcaataacaatccaagacaaagacaaat 2999  
|||||  
Db 2941 ttgccgatcaaatgtagccgaacatagaactgggcaataacaatccaagacaaagacaaat 3000  
|||||  
QY 3000 ccaacgctgcagcattataatgatataataatcacaggttttaacctataaaataataaca 3059  
|||||  
Db 3001 ccaacgctgcagcattataatgatataataatcacaggttttaacctataaaataataaca 3060  
|||||  
QY 3060 acccattgactttgtctccacttatgacattgttgactttggcaattggcaatggcaccac 3119  
|||||  
Db 3061 acccattgactttgtctccacttatgacattgttgactttggcaatggcaccac 3120  
|||||  
QY 3120 ccgccacagtaacccatgtatccgcttaacaaacccagtgtaagttgggtatgtatgtgaattg 3179  
|||||  
Db 3121 ccgccacagtaacccatgtatccgcttaacaaacccagtgtaagttgggtatgtatgtgaattg 3180  
|||||  
QY 3180 tggatgatacaacatttactaacaggcactgatgacataaaaaaacttggcgtcaaaa 3239  
|||||  
Db 3181 tggatgatacaacatttactaacaggcactgatgacataaaaaaacttggcgtcaaaa 3240  
|||||  
QY 3240 ccacaaactgaacaaacaaagtgttaattgtaatacagcaactaaacttttaattgtaact 3299  
|||||  
Db 3241 ccacaaactgaacaaacaaagtgttaattgtaatacagcaactaaacttttaattgtaact 3300  
|||||  
QY 3300 ctatgtatgaatgacctgtttaaagccaaacacatccgcaaaacttaaacacacctag 3359  
|||||  
Db 3301 ctatgtatgaatgacctgtttaaagccaaacacatccgcaaaacttaaacacacctag 3360  
|||||  
QY 3360 ccaaggaatttcaacaccacaaagggcagacagacccgcttacaacacctttaccggtta 3419  
|||||  
Db 3361 ccaaggaatttcaacaccacaaagggcagacagacccgcttacaacacctttaccggtta 3420  
|||||  
QY 3420 aaaaggttagatgaaaaataatgtctgtatgacgccaacgccaacacggtggtgcaaaaga 3479  
|||||  
Db 3421 aaaaggttagatgaaaaataatgtctgtatgacgccaacgccaacacggtggtgcaaaaga 3480  
|||||  
QY 3480 acgcaataatcaagttcaacaccttaacactcaaaagtgaacgggtcttaataattaaaa 3539  
|||||  
Db 3481 acgcaataatcaagttcaacaccttaacactcaaaagtgaacgggtcttaataattaaaa 3540  
|||||  
QY 3540 ccgcaaaaaattggttacggttacctttggcatttaacacacaaagcgggtcttaaaagccgga 3599  
|||||  
Db 3541 ccgcaaaaaattggttacggttacctttggcatttaacacacaaagcgggtcttaaaagccgga 3600  
|||||  
QY 3600 aaagcacaccttaaacgaaggtggtgttctatttaaaaaaacccactggttagcgaacaaatcc 3659  
|||||  
Db 3601 aaagcacaccttaaacgaaggtggtgttctatttaaaaaaacccactggttagcgaacaaatcc 3660  
|||||  
QY 3660 aagtcgggtgctgtatggcgtgagttggcgaaggttaataataatgggtgtgaggtgctg 3719  
|||||  
Db 3661 aagtcgggtgctgtatggcgtgagttggcgaaggttaataataatgggtgtgaggtgctg 3720  
|||||  
QY 3720 gcatgtatggcacaactgcattaccagagatgaaattggctttacttgggactaatggct 3779  
|||||  
Db 3721 gcatgtatggcacaactgcattaccagagatgaaattggctttacttgggactaatggct 3780  
|||||  
QY 3780 cacttgataaaagcaaacccaccttaagcaagcggttataacgaggttggtataaaga 3839  
|||||

Db 3781 cacttgataaaagcaaacccaccttaagcaagcggttataacgaggttggtataaaga 3840  
|||||  
QY 3840 ttaccaacattcaatcaggtgagattgccaaaacagacatgactgctgtgacagcgga 3899  
|||||  
Db 3841 ttaccaacattcaatcaggtgagattgccaaaacagacatgactgctgtgacagcgga 3900  
|||||  
QY 3900 agatttatgatttaaaaaacccgaaacttgaaaaaaatacagcagtagtgcacaaacagcac 3959  
|||||  
Db 3901 agatttatgatttaaaaaacccgaaacttgaaaaaaatacagcagtagtgcacaaacagcac 3960  
|||||  
QY 3960 aaaaactcttacacgaattctcagtagcagatgaacaaaggttaataactttacggttagta 4019  
|||||  
Db 3961 aaaaactcttacacgaattctcagtagcagatgaacaaaggttaataactttacggttagta 4020  
|||||  
QY 4020 acccttactccagttatgacacctcaaaagactctgtatgctatcaccttttcaggtgaaa 4079  
|||||  
Db 4021 acccttactccagttatgacacctcaaaagactctgtatgctatcaccttttcaggtgaaa 4080  
|||||  
QY 4080 acggcattaccacccaaggtataaaggtgtgtgctgtgtggcattgacaaaacccaag 4139  
|||||  
Db 4081 acggcattaccacccaaggtataaataaaggtgtgtgctgtgtggcattgacaaaacccaag 4140  
|||||  
QY 4140 gottaaacacgcttaagctgacgtgggttaataataatggcaaaagcattgttcattgaca 4199  
|||||  
Db 4141 gottaaacacgcttaagctgacgtgggttaataataatggcaaaagcattgttcattgaca 4200  
|||||  
QY 4200 gcaaaaaatggttcaaaaataccaacacagacttaagcaaacactctagcttaattgttaccattg 4259  
|||||  
Db 4201 gcaaaaaatggttcaaaaataccaacacagacttaagcaaacactctagcttaattgttaccattg 4260  
|||||  
QY 4260 ataaaggttagctacgacacacagacaggggcaataataatacaaaagcaagaacaaaccc 4319  
|||||  
Db 4261 ataaaggttagctacgacacacagacaggggcaataataatacaaaagcaagaacaaaccc 4320  
|||||  
QY 4320 gtcgcccagcattgttgatgtctaaagcgaggtcttaacttgcaggaacaaatggtgaag 4379  
|||||  
Db 4321 gtcgcccagcattgttgatgtctaaagcgaggtcttaacttgcaggaacaaatggtgaag 4380  
|||||  
QY 4380 cgggtgactttgtctccacttatgacaccttcaactttgcccagatggcaatgccaacaccc 4439  
|||||  
Db 4381 cgggtgactttgtctccacttatgacaccttcaactttgcccagatggcaatgccaacaccc 4440  
|||||  
QY 4440 ctaaggtgacttatgatgacacaaacagtaaaagtgctcttatgtatgctcaattgtgg 4499  
|||||  
Db 4441 ctaaggtgacttatgatgacacaaacagtaaaagtgctcttatgtatgctcaattgtgg 4500  
|||||  
QY 4500 atgatacaacattggaattaaagataaaaaaacttggcgttaaaaaacccaacatttgacca 4559  
|||||  
Db 4501 atgatacaacattggaattaaagataaaaaaacttggcgttaaaaaacccaacatttgacca 4560  
|||||  
QY 4560 gtactggcacaggtgctataaaaaatttgccttaagcaatcaagctactggcgatggcttg 4619  
|||||  
Db 4561 gtactggcacaggtgctataaaaaatttgccttaagcaatcaagctactggcgatggcttg 4620  
|||||  
QY 4620 tcaagcccagtgatatacgttgcctcatcaaaacacttatctggcgacatccaaactgcca 4679  
|||||  
Db 4621 tcaagcccagtgatatacgttgcctcatcaaaacacttatctggcgacatccaaactgcca 4680  
|||||  
QY 4680 aaggggcaagccaagcgaacactcagcaggtctatgtgagctggtgcaataaaggtca 4739  
|||||  
Db 4681 aaggggcaagccaagcgaacactcagcaggtctatgtgagctggtgcaataaaggtca 4740  
|||||  
QY 4740 tctatgacagtaacgatacaagctactatcaagcgaataatgatgcagcttgataaaa 4799  
|||||  
Db 4741 tctatgacagtaacgatacaagctactatcaagcgaataatgatgcagcttgataaaa 4800  
|||||  
QY 4800 ccaaaaagttgccaagcaaaactggtcgcccaagcccaaacccagatggcacattgg 4859  
|||||  
Db 4801 ccaaaaagttgccaagcaaaactggtcgcccaagcccaaacccagatggcacattgg 4860  
|||||  
QY 4860 ctaaaaatgagtgtaaatcagtcatttaacaaagcaaaagtaaatgatgcataaaagc 4919  
|||||

Db 4861 ctcaaatgaatgtcaaatcagtcattacaagaacaaagtaaatgatgccaaataaaagc 4920  
Qy 4920 aaggcatcaatgaagacaacgcctttgttaaggaacttgaaagcgctctctgataaca 4979  
Db 4921 aaggcatcaatgaagacaacgcctttgttaaggaacttgaaagcgctctctgataaca 4980  
Qy 4980 aaaccaaaccgcgcagtaacttggttgatttaaatgcggttgcccaaacacgcgtga 5039  
Db 4981 aaaccaaaccgcgcagtaacttggttgatttaaatgcggttgcccaaacacgcgtga 5040  
Qy 5040 cctttgcagggtatcacaggcaacgcgtcaaaaactggcgagactttgaccatcaag 5099  
Db 5041 cctttgcagggtatcacaggcaacgcgtcaaaaactggcgagactttgaccatcaag 5100  
Qy 5100 gtgggcaacagacacccaataagctaacaccgataataacatcggtgtgtagcaggtg 5159  
Db 5101 gtgggcaacagacacccaataagctaacaccgataataacatcggtgtgtagcaggtg 5160  
Qy 5160 atggtctcaactgtcaaaccttgccaaagacctaaaccaactcttaacagcgttaatgcaggtg 5219  
Db 5161 atggtctcaactgtcaaaccttgccaaagacctaaaccaactcttaacagcgttaatgcaggtg 5220  
Qy 5220 gcaccaaaatgatgacaaaagcggtgtctttgttagactcaagcggtcaagccaaaagcaa 5279  
Db 5221 gcaccaaaatgatgacaaaagcggtgtctttgttagactcaagcggtcaagccaaaagcaa 5280  
Qy 5280 acacccctgtctaaagtccaatggcctggacccctgggtggcaaggtcatcagtaattgg 5339  
Db 5281 acacccctgtctaaagtccaatggcctggacccctgggtggcaaggtcatcagtaattgg 5340  
Qy 5340 gcaaaaggcacaaaagataccga-cgctgccaatgtacaacagtttaaacgaagtagcgaac 5398  
Db 5341 gcaaaaggcacaaaagataccgaagcgtcccaatgtacacagtttaaacgaagtagcgaac 5400  
Qy 5399 ttgttggtctctgtcaatgcttgtaataatgaatacgcgtgcgcgaatacaggttaaaatgcc 5458  
Db 5401 ttgttggtctctgtcaatgcttgtaataatgaatacgcgtgcgcgaatacaggttaaaatgcc 5460  
Qy 5459 gacatcaaaaagaccocaaatcaggttctcatctatcaatcccgcaactgtcatcaaaagcagcc 5518  
Db 5461 gacatcaaaaagaccocaaatcaggttctcatctatcaatcccgcaactgtcatcaaaagcagcc 5520  
Qy 5519 acggtacttgccgttaaaagttaaataacgataccgaaacacttgccacttggttggtatacaa 5578  
Db 5521 acggtacttgccgttaaaagttaaataacgataccgaaacacttgccacttggttggtatacaa 5580  
Qy 5579 gtggcggttggtataaagacggaacgctaacgcgtattaaagcaatgtttggttgcaaaacc 5638  
Db 5581 gtggcggttggtataaagacggaacgctaacgcgtattaaagcaatgtttggttgcaaaacc 5640  
Qy 5639 caaaaagatggcagcaaaaagccctgctgcaccttaaacgcgcgcaggttcagaccacac 5698  
Db 5641 caaaaagatggcagcaaaaagccctgctgcaccttaaacgcgcgcaggttcagaccacac 5700  
Qy 5699 tatttgacaaacacccgcgagaagccattgacagaaataaatgaaacaggttatccgcttc 5758  
Db 5701 tatttgacaaacacccgcgagaagccattgacagaaataaatgaaacaggttatccgcttc 5760  
Qy 5759 ttccatgtcaacgatggcaataacagagcgtgtgttacaaggcggttaacgcgcatgtactca 5818  
Db 5761 ttccatgtcaacgatggcaataacagagcgtgtgttacaaggcggttaacgcgcatgtactca 5820  
Qy 5819 agtgcctcaggcaacgactcagtgccgataggtttccaggccaaagcgagatggtagaacc 5878  
Db 5821 agtgcctcaggcaacgactcagtgccgataggtttccaggccaaagcgagatggtagaacc 5880  
Qy 5879 gccgttgcatagggcagacaaacccaagcagggcaaccaatccatccgcctcgggtgataac 5938  
Db 5881 gccgttgcatagggcagacaaacccaagcagggcaaccaatccatccgcctcgggtgataac 5940  
Qy 5939 gcacaagccagggcgatcaatccatccgcctcgggtgataacggaaggtgagcaggtgaag 5998  
Db 5941 gcacaagccagggcgatcaatccatccgcctcgggtgataacggaaggtgagcaggtgaag 6000

Qy 5999 cactctggtgccatcgccgacccaagcactgttaaggctgataacagttacagttggtt 6058  
Db 6001 cactctggtgccatcgccgacccaagcactgttaaggctgataacagttacagttggtt 6060  
Qy 6059 aataacaacaacagtttaccgatgccactcaaacacgatgtctttggttggtggaataacatc 6118  
Db 6061 aataacaacaacagtttaccgatgccactcaaacacgatgtctttggttggtggaataacatc 6120  
Qy 6119 accgtgacccgaagttaactcggttgctttaggttcaaaactctgccaactcagtgacggcaca 6178  
Db 6121 accgtgacccgaagttaactcggttgctttaggttcaaaactctgccaactcagtgacggcaca 6180  
Qy 6179 cagcagggcacacaaagcccaaaaatctgacggcacagcagggtacacaccacacagcaggt 6238  
Db 6181 cagcagggcacacaaagcccaaaaatctgacggcacagcagggtacacaccacacacagcaggt 6240  
Qy 6239 gcaacccgtacagggttaaaaggctttgttggaacaaacgcgggttggtgaggtctccgtggtt 6298  
Db 6241 gcaacccgtacagggttaaaaggctttgttggaacaaacgcgggttggtgaggtctccgtggtt 6300  
Qy 6299 gcctcaggtgctgaacgcgctatccaatgtggaacacgcgggttggtgaggtctccgtggtt 6358  
Db 6301 gcctcaggtgctgaacgcgctatccaatgtggaacacgcgggttggtgaggtctccgtggtt 6360  
Qy 6359 accgatgcggttcaatggttagccagttgtacaaagccacccaagcattgccaacgcaacc 6418  
Db 6361 accgatgcggttcaatggttagccagttgtacaaagccacccaagcattgccaacgcaacc 6420  
Qy 6419 aatgagcttgaccatcgctatccacaaacgaataaaggccaaatgcagggtattcaca 6478  
Db 6421 aatgagcttgaccatcgctatccacaaacgaataaaggccaaatgcagggtattcaca 6480  
Qy 6479 gcgatggcgatggggtccatgcacaaacgctcacatctctgcagatccatggtttaccggg 6538  
Db 6481 gcgatggcgatggggtccatgcacaaacgctcacatctctgcagatccatggtttaccggg 6540  
Qy 6539 ggtattgccaacccacacaggttcaagggtgcggtggcagtggtgactgtcgaagctgtcgat 6598  
Db 6541 ggtattgccaacccacacaggttcaagggtgcggtggcagtggtgactgtcgaagctgtcgat 6600  
Qy 6599 aatggtcaatgggtatttaaaatcaatggttcagcgatatacccaaggccatgtaggggcg 6658  
Db 6601 aatggtcaatgggtatttaaaatcaatggttcagcgatatacccaaggccatgtaggggcg 6660  
Qy 6659 gcagttggtgcaggttttcaacttttaagccataaatacgaagattttacttaaaatacaa 6718  
Db 6661 gcagttggtgcaggttttcaacttttaagccataaatacgaagattttacttaaaatacaa 6720  
Qy 6719 tctcaccatagttgtat-aaaaacgatacagatcagtcacatattactgatgctgatttt 6777  
Db 6721 tctcaccatagttgtataaaaaacgatacagatcagtcacatattactgatgctgatttt 6780  
Qy 6778 tttatcaacttaaacattttaccgctcaagtgattctctttcacccatgacaaaatcgcca 6837  
Db 6781 tttatcaacttaaacattttaccgctcaagtgattctctttcacccatgacaaaatcgcca 6840  
Qy 6838 ttgatcataggttaaaacttatgtagtaaaattttatcaatgtagttgttagatatgttaaa 6897  
Db 6841 ttgatcataggttaaaacttatgtagtaaaattttatcaatgtagttgttagatatgttaaa 6900  
Qy 6898 atttgccattgacaaaaaaatgaccgattttatcccgaaaaatttctgattgatccggtt 6957  
Db 6901 atttgccattgacaaaaaaatgaccgattttatcccgaaaaatttctgattgatccggtt 6960  
Qy 6958 gacctgcaggtcgac 6972  
Db 6961 gacctgcaggtcgac 6975

RESULT 13

US-09-361-619-12

; Sequence 12, Application US/09361619



: GENERAL INFORMATION:
: APPLICANT: Loosmore, Sheena M.
: APPLICANT: Sasaki, Ken
: APPLICANT: Yang, Yan Ping
: APPLICANT: Klein, Michel H.
: TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
: FILE REFERENCE: 1038-921MIS:jb
: CURRENT APPLICATION NUMBER: US/09/361, 619
: CURRENT FILING DATE: 1999-07-27
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 12
: LENGTH: 6259
: TYPE: DNA
: ORGANISM: Moraxella catarrhalis
: us-09-361-619-12

Query Match	89.8%;	Score 6258;	DB 17;	Length 6259;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 6258;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 708	tgatcggtgcaacgctcagtggtgagtcgttctgctcaaaaaaagataccaaacatctcg	767		
DB 2	tgatcggtgcaacgctcagtggtgagtcgttctgctcaaaaaaagataccaaacatctcg	61		
QY 768	caattgtgaaacaaacagcagcgtcagcagcgtccaaaggcgagcggtgacgcag	827		
DB 62	caattgtgaaacaaacagcagcgtcagcagcgtccaaaggcgagcggtgacgcag	121		
QY 828	ccattgctattgtgaaatgctaaacgacagcggtcgaagccatcgccatcggtagta	887		
DB 122	ccaatgctattgtgaaatgctaaacgacagcggtcgaagccatcgccatcggtagta	181		
QY 888	gtaataaactgtcaatggaagcagtttggataagatagtagtaaccgatgctacgggtcgaag	947		
DB 182	gtaataaactgtcaatggaagcagtttggataagatagtagtaaccgatgctacgggtcgaag	241		
QY 948	agtcacatgcacatcggttggtgagatgaagcagtagtgatgctcgatgctcgatgctcggtta	1007		
DB 242	agtcacatgcacatcggttggtgagatgaagcagtagtgatgctcgatgctcgatgctcggtta	301		
QY 1008	gtgatgacttacattgtctgtatcagcagatgtaacatcctaaccatccgaaggtactctga	1067		
DB 302	gtgatgacttacattgtctgtatcagcagatgtaacatcctaaccatccgaaggtactctga	361		
QY 1068	ttacagatcttattaaacggccatgcagtagtattaaagaatacgaagctcaaaagataatg	1127		
DB 362	ttacagatcttattaaacggccatgcagtagtattaaagaatacgaagctcaaaagataatg	421		
QY 1128	atgtataaataatagacgcacacccgaagcgacacgcagtcactgagtgaggagccatgt	1187		
DB 422	atgtataaataatagacgcacacccgaagcgacacgcagtcactgagtgaggagccatgt	481		
QY 1188	catatgcacagggctattttccaaacgcttggtagcaggggcaacgctcaaaagtgccct	1247		
DB 482	catatgcacagggctattttccaaacgcttggtagcaggggcaacgctcaaaagtgccct	541		
QY 1248	attccttgagcagtggtcttccgcacacgcagcgagggccaaatctacaatcgctattggtt	1307		
DB 542	attccttgagcagtggtcttccgcacacgcagcgagggccaaatctacaatcgctattggtt	601		
QY 1308	ctgatgcaacatctagctcgtttggagcgatagcccttggtagcaggtactcgtctcagc	1367		
DB 602	ctgatgcaacatctagctcgtttggagcgatagcccttggtagcaggtactcgtctcagc	661		
QY 1368	tacagggcagattgctcagtagtcaaggtctgttctgctcactcagatgataataattcta	1427		
DB 662	tacagggcagattgctcagtagtcaaggtctgttctgctcactcagatgataataattcta	721		
QY 1428	gaccggcctatcacacaaatcccaggcactagaccccaagtttcaagcccaacataata	1487		

DB 722	gaccggcctatcacacaaataccaggcactagaccccaagtttcaaggcccaacataata	781		
QY 1488	cgaaggcgggtccactttccattcgttagtaactctatacaacgtataaatacatcaatgtcg	1547		
DB 782	cgaaggcgggtccactttccattcgttagtaactctatacaacgtataaatacatcaatgtcg	841		
QY 1548	gtgcaggtgttaataaaacccgagtcggtcaatgtggtcacagctagaagcggtggtgaagt	1607		
DB 842	gtgcaggtgttaataaaacccgagtcggtcaatgtggtcacagctagaagcggtggtgaagt	901		
QY 1608	gggtcaaggagcgtgagatattactttcaggtgagtaacacagctactgacgtataaataag	1667		
DB 902	gggtcaaggagcgtgagatattactttcaggtgagtaacacagctactgacgtataaataag	961		
QY 1668	gtttggataatacttttaactattaaagggtggtcgtgaaagttaaaactgtcaaaactt	1727		
DB 962	gtttggataatacttttaactattaaagggtggtcgtgaaagttaaaactgtcaaaactt	1021		
QY 1728	atateggtgtgttaaaagagcgtgataatagtggtcgtgaaagttaaaactgtcaaaactt	1787		
DB 1022	atateggtgtgttaaaagagcgtgataatagtggtcgtgaaagttaaaactgtcaaaactt	1081		
QY 1788	taacaacatcttactgagtggaatacaactacattataatgccacaacccacagtttaaggtag	1847		
DB 1082	taacaacatcttactgagtggaatacaactacattataatgccacaacccacagtttaaggtag	1141		
QY 1848	gtagtagtagtagtactacagctgaattattgagtgatagtttaacctttaccacagccca	1907		
DB 1142	gtagtagtagtagtactacagctgaattattgagtgatagtttaacctttaccacagccca	1201		
QY 1908	atacaggcagtcgaagcacaagcacaacccgtctatggtcgttgaatgggtgagtttacta	1967		
DB 1202	atacaggcagtcgaagcacaagcacaacccgtctatggtcgttgaatgggtgagtttacta	1261		
QY 1968	ataatgcgaacaaacagcagcaatcggcactactcgtattaccagagataaaaaatggct	2027		
DB 1262	ataatgcgaacaaacagcagcaatcggcactactcgtattaccagagataaaaaatggct	1321		
QY 2028	ttgctcgagatggtgatgttgatgaaaaaacaagcaccatatttggataaaaaaacactta	2087		
DB 1322	ttgctcgagatggtgatgttgatgaaaaaacaagcaccatatttggataaaaaaacactta	1381		
QY 2088	aaagtggtagtgtgcaattaccatagacaaatggcatttgatgacgttaataaaaaagatca	2147		
DB 1382	aaagtggtagtgtgcaattaccatagacaaatggcatttgatgacgttaataaaaaagatca	1441		
QY 2148	gtaactctgccaaaggtagcagtgctaaacgctgcggttaccatcgaaacagctcaaaagccg	2207		
DB 1442	gtaactctgccaaaggtagcagtgctaaacgctgcggttaccatcgaaacagctcaaaagccg	1501		
QY 2208	ccaagcctacttttaaacgcagcgctggcctcagtgctcacctactgaaatcatcagttg	2267		
DB 1502	ccaagcctacttttaaacgcagcgctggcctcagtgctcacctactgaaatcatcagttg	1561		
QY 2268	atgctaagagtggtgcaatgtttaccgcccccaacttacaacattggcgtgaaaaaccacgagc	2327		
DB 1562	atgctaagagtggtgcaatgtttaccgcccccaacttacaacattggcgtgaaaaaccacgagc	1621		
QY 2328	ttaacagtgatgcacatagtgataaaatttagtgttaagggttagtgtagcgaacaaatagct	2387		
DB 1622	ttaacagtgatgcacatagtgataaaatttagtgttaagggttagtgtagcgaacaaatagct	1681		
QY 2388	tagttaccgcccgaacatttggcgaagctctataatgaagtcaatcgaaagcgtcgacagtg	2447		
DB 1682	tagttaccgcccgaacatttggcgaagctctataatgaagtcaatcgaaagcgtcgacagtg	1741		
QY 2448	ctctacaaagctttacacgtttaaagaagaagcagatgacgccaacgctatcacccgttg	2507		
DB 1742	ctctacaaagctttacacgtttaaagaagaagcagatgacgccaacgctatcacccgttg	1801		
QY 2508	ctaaagatagcacaacaaatgcgcggcgagctcagcatcttataaactcaaaaggtataaacg	2567		
DB 1802	ctaaagatagcacaacaaatgcgcggcgagctcagcatcttataaactcaaaaggtataaacg	1861		









Db 1681 aacgcaggcgtgcatcagtgatcacaactactgaaatataatcatggtgatgctaaagtggtgc 1740  
Qy 2282 aatgtttaccgcccacattacaacattgctgtaaaacacacacagcttaaacagtgatgac 2341  
Db 1741 aatgtttaccgcccacattacaacattgctgtaaaacacacacagcttaaacagtgatgac 1800  
Qy 2342 actagtgataaaattagtgtaaggttagtgtagaacaataagcttagttaccgcccga 2401  
Db 1801 actagtgataaaattagtgtaaggttagtgtagaacaataagcttagttaccgcccga 1860  
Qy 2402 catttggaagctatatataatgaagtcacatgaacagcgtgacagtgctctacaagaagcttt 2461  
Db 1861 catttggaagctatatataatgaagtcacatgaacagcgtgacagtgctctacaagaagcttt 1920  
Qy 2462 accgtttaaagaagaagcagatgacgcacacgctatcacgctggtctaaagatagcaca 2521  
Db 1921 accgtttaaagaagaagcagatgacgcacacgctatcacgctggtctaaagatagcaca 1980  
Qy 2522 aaaaatgcggcgagtcagcatctttaaactcaaaaggttaaaacggtctaacggttgct 2581  
Db 1981 aaaaatgcggcgagtcagcatctttaaactcaaaaggttaaaacggtctaacggttgct 2040  
Qy 2582 accaaaaagatggtacggttacccttggtctagccaagatagcgtctgaccattgac 2641  
Db 2041 accaaaaagatggtacggttacccttggtctagccaagatagcgtctgaccattgac 2100  
Qy 2642 aaaaacacccataaacacagctggttgactgttaaagatagccaagcgaacaaatccaagtc 2701  
Db 2101 aaaaacacccataaacacagctggttgactgttaaagatagccaagcgaacaaatccaagtc 2160  
Qy 2702 ggtgctaaatggcattaaattactaatgtgaatggtagtaataccaggtactggcattgca 2761  
Db 2161 ggtgctaaatggcattaaattactaatgtgaatggtagtaataccaggtactggcattgca 2220  
Qy 2762 aatcacgctgcattaccagagataaaattgcttggcttctgagtggtgagtgat 2821  
Db 2221 aatcacgctgcattaccagagataaaattgcttggcttctgagtggtgagtgat 2280  
Qy 2822 acaaaacaaacttatcttgatcaagaacagctacaagttggcaatgttaagattaccaac 2881  
Db 2281 acaaaacaaacttatcttgatcaagaacagctacaagttggcaatgttaagattaccaac 2340  
Qy 2882 actggcattaaacgaggtgtaaaagccatcaacaggtgctcccaacactgctacgatt 2941  
Db 2341 actggcattaaacgaggtgtaaaagccatcaacaggtgctcccaacactgctacgatt 2400  
Qy 2942 gccgataaagtagcgcaacatagaaactgggcaatacaaatccaaagacaaagacaaatcc 3001  
Db 2401 gccgataaagtagcgcaacatagaaactgggcaatacaaatccaaagacaaagacaaatcc 2460  
Qy 3002 aacgctgcccgcatatgatataataataacaggttttaacctaaataataatacaac 3061  
Db 2461 aacgctgcccgcatatgatataataataacaggttttaacctaaataataatacaac 2520  
Qy 3062 ccaattgaacttgcctccacttatgacattggtgactttgcaatggcaatgcccacac 3121  
Db 2521 ccaattgaacttgcctccacttatgacattggtgactttgcaatggcaatgcccacac 2580  
Qy 3122 gccacagtaaccatgataccgcttaacaaacacagtaaaagtgttatgtatgtgaaatgtg 3181  
Db 2581 gccacagtaaccatgataccgcttaacaaacacagtaaaagtgttatgtatgtgaaatgtg 2640  
Qy 3182 gatgatacaaccattatcaacaggtcactgatgacataataaaacttggcgtcaaaac 3241  
Db 2641 gatgatacaaccattatcaacaggtcactgatgacataataaaacttggcgtcaaaac 2700  
Qy 3242 accaaactgaacaaaacaaagtgttaaattggttaatacacagcaactaaactttaaactct 3301  
Db 2701 accaaactgaacaaaacaaagtgttaaattggttaatacacagcaactaaactttaaactct 2760  
Qy 3302 agtgatgaagatgcccttgttaacgccaagacatcgccgaaataatctaaacacccatgac 3361  
Db 2761 agtgatgaagatgcccttgttaacgccaagacatcgccgaaataatctaaacacccatgac 2820

Qy 3362 aaggaaattcacaccacaaaggcacacagcagacacccgctctacaaacottttaccgttaaa 3421  
Db 2821 aaggaaattcacaccacaaaggcacacagcagacacccgctctacaaacottttaccgttaaa 2880  
Qy 3422 aaggtagatgaaaaataaataatgctgtagtacgccaacgcatcacgctgggtcacaagaac 3481  
Db 2881 aaggtagatgaaaaataaataatgctgtagtacgccaacgcatcacgctgggtcacaagaac 2940  
Qy 3482 gcaataatcaagtcacacacccctaacactcaaaaggtgaaaacggtcttaataattaaaacc 3541  
Db 2941 gcaataatcaagtcacacacccctaacactcaaaaggtgaaaacggtcttaataattaaaacc 3000  
Qy 3542 gacaaaaatggttaoagttacottttggcattaacacacacagcgtcttaaaagccgcaaa 3601  
Db 3001 gacaaaaatggttaoagttacottttggcattaacacacacagcgtcttaaaagccgcaaa 3060  
Qy 3602 agcacccataacacagcgttggcttcttaataaaacccactggtagcgaacaaatccaa 3661  
Db 3061 agcacccataacacagcgttggcttcttaataaaacccactggtagcgaacaaatccaa 3120  
Qy 3662 gtcggtgctgtagggcgtgaaagtttgcgaaggttaataataatgggtttaggtgctggc 3721  
Db 3121 gtcggtgctgtagggcgtgaaagtttgcgaaggttaataataatgggtttaggtgctggc 3180  
Qy 3722 attgtagcacaactcgcatcattaccagagatgaaattggctttacttgggactaaatgctca 3781  
Db 3181 attgtagcacaactcgcatcattaccagagatgaaattggctttacttgggactaaatgctca 3240  
Qy 3782 ctgtataaaagcaaacccacacttaagcaagcagcatttaacgcaggtgggttaaaagatt 3841  
Db 3241 ctgtataaaagcaaacccacacttaagcaagcagcatttaacgcaggtgggttaaaagatt 3300  
Qy 3842 accaacttaacatcaggtgagatttgcacaaacagccatgatacgtgtagcaggtggcgaag 3901  
Db 3301 accaacttaacatcaggtgagatttgcacaaacagccatgatacgtgtagcaggtggcgaag 3360  
Qy 3902 atttatgatttaaaacagcagcattgaaacaaatcagcagctacgcaaaacagacaa 3961  
Db 3361 atttatgatttaaaacagcagcattgaaacaaatcagcagctacgcaaaacagacaa 3420  
Qy 3962 aactcattacacgaattctcagtagcagatgaacaaaggttaataactttacggttagtaac 4021  
Db 3421 aactcattacacgaattctcagtagcagatgaacaaaggttaataactttacggttagtaac 3480  
Qy 4022 cttactcagttatgacacccctcaaaagcctctgtagtcatcacctttgaggtgaaac 4081  
Db 3481 cttactcagttatgacacccctcaaaagcctctgtagtcatcacctttgaggtgaaac 3540  
Qy 4082 ggcattaccacaaaggttaaaaggttgggtgctggtggcattgacacaaacaaagc 4141  
Db 3541 ggcattaccacaaaggttaaaaggttgggtgctggtggcattgacacaaacaaagc 3600  
Qy 4142 ttaaccacgcctaaagctgacggtgggtgtaataataatggcaaggttgcattgacagc 4201  
Db 3601 ttaaccacgcctaaagctgacggtgggtgtaataataatggcaaggttgcattgacagc 3660  
Qy 4202 caaaaatgggtcaaaaataccatcagagcagcagcagcagcagcagcagcagcagcagcagc 4261  
Db 3661 caaaaatgggtcaaaaataccatcagagcagcagcagcagcagcagcagcagcagcagcagc 3720  
Qy 4262 aaaggtagcgtacgcacacacagcagcagcagcagcagcagcagcagcagcagcagcagc 4321  
Db 3721 aaaggtagcgtacgcacacacagcagcagcagcagcagcagcagcagcagcagcagcagc 3780  
Qy 4322 gccgcagc 4381  
Db 3781 gccgcagc 3840  
Qy 4382 gttgacttgcctccacttatgacacccgctcaacttggcagatggcagcagcagcagcagcagc 4441  
Db 3841 gttgacttgcctccacttatgacacccgctcaacttggcagatggcagcagcagcagcagcagc 3900

Db	4981	gtacttggcgttaaggttaataacgataccgaaacattgccactgggtatatacagtg	5040
Qy	5582	ggcgttgataaagacgcgaacactaacgcgattaaagcaaatgtttgggtcaaaacccaa	5641
Db	5041	ggcgttgataaagacgcgcacgttaacgcgattaaagcaaatgtttgggtcaaaacccaa	5100
Qy	5042	aaagatggcagcaaaaaagccctgctgcacactataacgcgcgaggtcgagaccaactat	5701
Db	5101	aaagatggcagcaaaaaagccctgctgcacactataacgcgcgaggtcgagaccaactat	5160
Qy	5702	ttgaccaaaaccccgacgaagccattgacagataataatgaacaaagtataccgtttcttc	5761
Db	5161	ttgaccaaaaccccgacgaagccattgacagataataatgaacaaagtataccgtttcttc	5220
Qy	5762	catgtcaacgatggcaatcaagagcctgtgtgtacaaaggcgttaacgcgattgactcaagt	5821
Db	5221	catgtcaacgatggcaatcaagagcctgtgtgtacaaaggcgttaacgcgattgactcaagt	5280
Qy	5822	gcctcaggcgaagcactcagtggcgtataggtttccaggccaaaggcagatgggtgaagccgcc	5881
Db	5281	gcctcaggcgaagcactcagtggcgtataggtttccaggccaaaggcagatgggtgaagccgcc	5340
Qy	5882	gttgccataggcagacaaacccaagcaggcaaccaatccatgcgcactcgtgtgataacgcga	5941
Db	5341	gttgccataggcagacaaacccaagcaggcaaccaatccatgcgcactcgtgtgataacgcga	5400
Qy	5942	caagccacgggcgatcaatccatgccactcgttacaggcaatgtgttagcaggttaagcac	6001
Db	5401	caagccacgggcgatcaatccatgccactcgttacaggcaatgtgttagcaggttaagcac	5460
Qy	6002	tctgtgtccactcggcgaccacaaagcactgttaaggctgataaacagttacagttgtgggtaat	6061
Db	5461	tctgtgtccactcggcgaccacaaagcactgttaaggctgataaacagttacagttgtgggtaat	5520
Qy	6062	aacaaacagtttacccgatgccactcaaaacccgatgtcttgggtgggccaataaacatcacc	6121
Db	5521	aacaaacagtttacccgatgccactcaaaacccgatgtcttgggtgggccaataaacatcacc	5580
Qy	6122	gtgacgcgaagaataactcgggttgccttaggttcaaaactcgcgcactcagtgacggaacacac	6181
Db	5581	gtgacgcgaagaataactcgggttgccttaggttcaaaactcgcgcactcagtgacggaacacac	5640
Qy	6182	gcagggcacacaaagcccaaaaaattcgacgcacagcaggttaacaccacacagcaggtgca	6241
Db	5641	gcagggcacacaaagcccaaaaaattcgacgcacagcaggttaacaccacacagcaggtgca	5700
Qy	6242	accggtacgggttaaagcgtttcgtggacaaaacggcggttgggtgcggttcctcgtgggtgcc	6301
Db	5701	accggtacgggttaaagcgtttcgtggacaaaacggcggttgggtgcggttcctcgtgggtgcc	5760
Qy	6302	tcaggtctgaacgcgcgtatccaaaatgtggcagcaggtgaggtcagtgccaccagcacc	6361
Db	5761	tcaggtctgaacgcgcgtatccaaaatgtggcagcaggtgaggtcagtgccaccagcacc	5820
Qy	6362	gatcggtcgaattagtcaggttgtacaagccacccaaagcatttgccaaacgcgaaccaat	6421
Db	5821	gatcggtcgaattagtcaggttgtacaagccacccaaagcatttgccaaacgcgaaccaat	5880
Qy	6422	gaggttgaccatcgtatccaccacaaaacgaaaataaggccaatgcagggtatttcacgag	6481
Db	5881	gaggttgaccatcgtatccaccacaaaacgaaaataaggccaatgcagggtatttcacgag	5940
Qy	6482	atggcgatggcgtccatgccacaagcctacattcctggcagatccatggttaccggggggt	6541
Db	5941	atggcgatggcgtccatgccacaagcctacattcctggcagatccatggttaccggggggt	6000
Qy	6542	attgccaccacaaacggtcaaggtgcggtggcagtggaagctgtcgaaagctgtcggaataat	6601
Db	6001	attgccaccacaaacggtcaaggtgcggtggcagtggaagctgtcgaaagctgtcggaataat	6060
Qy	6602	gggtcaatgggtatttaaaatcaatggttcagccgatacccaaggccatgtgagggcgcca	6661

Db 6061 ggtcaatgggtatttaaaatcaatggttcaagccgatacccaaggccatgtagggcgga 6120

Qy 6662 gttgtgcaggttttcaacttt 6682

Db 6121 gttgtgcaggttttcaacttt 6141

RESULT 15

US-08-621-944-2

; Sequence 2, Application US/08621944

; GENERAL INFORMATION:

; APPLICANT: SASAKI, Ken

; APPLICANT: HARKNESS, Robin E.

; APPLICANT: LOOSMORE, Sheena M.

; APPLICANT: KLEIN, Michel H.

; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER

; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/621,944

; FILING DATE: 26-MAR-1996

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/478,370

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-587

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5976 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-621-944-2

Query Match 85.7%; Score 5976; DB 10; Length 5976;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 5976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 707 gtgatcggtgcaacgctcagtgccagtgcttctcaaaaaaaagatacccaacatc 766

Db 1 GTGATCGGTGCAACGCTCAGTGCGCAGTGCTATGCTCAAAAAAAGATACCAACATATC 60

Qy 767 gcaattggtgaacaaaaccagccagcgtcagggcactgccaaaggcgagcggtgatcga 826

Db 61 GCAATTGCTGAACAAAACCAAGCAAGCGCTCAGGCACTGCCAAGGCGGACGGTGATCGA 120

Qy 827 gccattcctattgtgaaatgctaaacgacagcggtcaagccatcgccatcggttagt 886

Db 121 GCCATTGCTATTGTTGAAATGCTAAACGACAGCGCGGTCAAGCCATCGCCATCGGTAGT 180

Qy 887 agtaataaaactgtcaatggagcagtttggatagaagttagtaccgtatgtaaggggtcaa 946

Db 181 AGTAATAAACTGTCAATGGAAGCAGTTGGATAAGATAGGTACCGATGTACGGGTCAA 240

Qy 947 gagtccatcgccatcggtggtgataaaaggctagtggtgattgctcctcgattgccatcggt 1006

Db 241 GAGTCCATCGCCATCGTGTGATGTAAGGCTAGTGTGATGCTCGATTGCCATCGGT 300

Qy 1007 agtgaacttacatttctggtgatacagcagtggttaactcctaataacatccgaagaagctctg 1066

Db 301 AGTGATGACTTACATTTGCTTGATTCAGCATGGTAATCCTTAACATCCGAAGGTACTCTG 360

Qy 1067 attaacgatcttattaaagccatcgagttattaaagaatacgaagactcaaaaggataat 1126

Db 361 ATTAAACGATCTTATTAAAGCCATGTCAGTATTAAAGAATATACGAAGCTCAAGGATAAT 420

Qy 1127 gatgtaaaatatagacgcacaaacgcgaagcgagccacccagtcagtgagtgagccatg 1186

Db 421 GATGTAATAATATAGACGCACAAACGCAAGCGGACACGCCAGTACTGCAGTGGGAGCCATG 480

Qy 1187 tcatacgacagggctatttttccaaagcctttgttacacgggcaacagctcaaaagtcgc 1246

Db 481 TCATATGCACAGGGTCATTTTTCACAGCCCTTTGGTACACGGGCAACAGCTTAAAGTGCC 540

Qy 1247 tattccttggcagtggttcttgcgcgcacacagcgagggcccaatctacaatcgctattggt 1306

Db 541 TATTCTTGGCAGTGGGTCTTGGCCGCACACGCCGAGGGCCCAATCTACAATCGCTATTGGT 600

Qy 1307 tctgatgcaacatctagctcgttggggagcgatagcccttggcgaggtactcgtgctcag 1366

Db 601 TCTGATGCAACATCTAGTCTCGTTGGGAGCGATAGCCCTTGGTGCAGGTACTCGTGTCTCAG 660

Qy 1367 ctacagggcagttatgcccctaggttcaaggttctgtgtcactcagagtgataaattct 1426

Db 661 CTACAGGGCAGTATTGGCCCTAGGTCAAGTTCTGTTGTCACCTCAGAGTGATTAATTTCT 720

Qy 1427 agaccggcctatacaccaaaataccagggcactagaccaccaagtttcaagccaccaataat 1486

Db 721 AGACCGGCCTATACACCAAAATACCCAGGCACCTAGACCCCAAGTTTCAAGCCACCAATAAT 780

Qy 1487 acgaaggcggttccactttccattggttagtaactctcaaacgttaaatcatcaatc 1546

Db 781 ACGAAGCGGGTCCACTTTTCCATTGCTAGTAACTCTATCAAACTTAAATCATCATGTC 840

Qy 1547 ggtgcaggtgttaataaaacccagtcggtcgaatgtggcacagctagaagcggtggtgaag 1606

Db 841 GGTGCAGGTGTTTAATAAAACCGATCGGTCAATGTGGCACAGCTAGAAAGCGGTGTGAAG 900

Qy 1607 tgggctaaggcgctagaaattacttttcagggtgatacagctacgtacgtacgtacgtacgtac 1666

Db 901 TGGGCTAAGGAGCGTAGAATTACTTTTCAGGGGTGATGATAACAGTACTGACGTAAATA 960

Qy 1667 ggttggataataacttacttaactataaagtggtgcagagaccaacgcattaccgcataat 1726

Db 961 GGTGTTGCATAAATCTTTAACTTAAAGTGTGTCAGAGACCAACGCATTAACCCGATAAT 1020

Qy 1727 aatacgggtgtgttaaaagagcgctgataatagtggtctgaaagtttaaaacttaccgaccc 1786

Db 1021 AATATCGGTGTGTAAGAGAGGCTGATATAGTGTGCTGTAAGTTTAAACTTCTGCTAAACT 1080

Qy 1787 ttaacaaacttactgaggtgatacaactacattaaatgcccacacacacagcttaaggtca 1846

Db 1081 TTAACAACTTCTTACTGAGGTGATACACTATTAATGCCCAACACACACGTTAAGGTA 1140

Qy 1847 ggtagtagtagtagtactacagctgaattattgagtgatagtttaacctttaccgacccc 1906

Db 1141 GGTAGTAGTAGTAGTACTACAGCTGAATTTATTGAGTGATAGTTTAACTTTTACCAGGCC 1200

Qy 1907 aatacaggcgagtcgaagcgaacgaacacccgctctatggcgttaagtgggtgagtttact 1966

Db 1201 AATACAGGAGGCTCAAGAGCAAAACCGCTATGCGGTGTTAATGGGTGAAGTTTACT 1260

Qy 1967 aataatgcagaacacacagcagcaatcgccactactcgttaccagagagataaaattggc 2026

Db 1261 AATAATGCAGAAACACACAGCAGCAATCGGCACCTACTCTGTTATTACCAGAGATATAAATTGGC 1320



QY 2027 ttgtgtcagatggtgtagtgaataaagcaaacacacatttttgataaaaaaacatt 2086  
DB 1321 TTTGTCGAGATGGTGTGATGATAAAGCAAGCACCATTATTTGGATAAAAAACAATT 1380  
QY 2087 aaagtgggtagtggtgcaattaccatagacaaatggcattgatgagcaggttaataaaagatc 2146  
DB 1381 AAAGTGGGTAGTGTCCAAATTAACCATACAAATGGCAATGTGACAGGTAAATAAAGATC 1440  
QY 2147 agtaacttgcgaaggtagcagtgctaacgatgcggtttacatcgaacagctcaaaagcc 2206  
DB 1441 AGTAATCTTGCCAAAGGTAGCAGTGCTAACGATGGGTTACCATGGAACAGCTCAAGCC 1500  
QY 2207 gccaaagctactttaaacgagggcgctggcatcagtgcaacactactgaataatcagtt 2266  
DB 1501 GCCAAGCCTACTTTAAACGCGAGCGCTGGCATCAGTGTGCACACCTTACTGAAATATACATT 1560  
QY 2267 gatcgaagagtggaattgttacgcgcacacacttaacacattggcgtgaataaacacagag 2326  
DB 1561 GATGCTAAGAGTGGCAATGTTACCGCCCACTTACACATTTGGGTTGAAACACCGAG 1620  
QY 2327 cttaacagtgatggcactagtataaatttagtgttaagggttagtgggtacgaacaatagc 2386  
DB 1621 CTTAACAGTGATGGCAGTGTGATAAATTTAGTGTAAAGGTAGTGGTACGAACAATAGC 1680  
QY 2387 ttagtaccgcggaacatttggcaagctatctaaatgaagtcacatcgcaacggctgacagt 2446  
DB 1681 TTAGTTACCGCGGAACATTTGGCAAGCTATCTAAATGAAGTCAATCGAAGCGCTGACAGT 1740  
QY 2447 gctcacaagcttaccgtttaaagaagaagcagatgatgcgcgaacgcctatcacctg 2506  
DB 1741 GCTCTACAAAGCTTTACCGTTAAAGAAAGACGATGATGACCGCAACGCTATCACCGTG 1800  
QY 2507 gctaaagatacgaacaaaaatgcccgcgcagtcagcatcttaaaactcaaaaggttaaaac 2566  
DB 1801 GCTAAAGATACGACAAAAATGCGCGCGCAGTCAGCATCTTAAACTCAAAGGTAAAAAC 1860  
QY 2567 ggtcgaaggttgctacaaaaaaagatggtacggttaacctttgggcttagccaagatagc 2626  
DB 1861 GGTCTAACCGGTGTCTACCAAAAAAGATGGTACGGTTAGCTTTGGGCTTAGCCAAAGTAGC 1920  
QY 2627 ggtcgcacattggcaaacgacccctaaacacagatggtgactggttaagatcaccaac 2686  
DB 1921 GGTCTGACCATTTGGCAAAAGCACCTTAACACAGATGCTTGACTGTTAAAGATACCAAC 1980  
QY 2687 gaacaaaatccagtcggtgcttaattggcattaaattactaatgtgaatggtagtaacca 2746  
DB 1981 GAACAAATCCAAAGTCGGTGCTAATGGCATTAATTTACTAATGTGAATGGTAGTAATCCA 2040  
QY 2747 ggtactggcattgcaaatccgcctgcgaattaccagagataaaattggctttgctggtct 2806  
DB 2041 GGTACTGGCATTTGCAAAATACCGCTCGCATTACCAGAGATAAAATTTGGCTTTGCTGGTTCT 2100  
QY 2807 gatggtgcagttgatacaaaaacacttatcttgatcaagacaagctacaagttggcaat 2866  
DB 2101 GATGCTGCAGTTGATACAAACAACTTATCTTGTATCAAGACAACTACAAGTTGGCAAT 2160  
QY 2867 gttaaagattaccaacactggcatcaacgcaggtggttaagccatcacagggtgtcccca 2926  
DB 2161 GTTAAGATTACCAACACTGGCATTAACGCAGGTGTTAAAGCCATCACAGGGGTGTCCCCA 2220  
QY 2927 acactgcctagattgcgcatcaaaagtgcgcgaacacatagaaactgggcaatacaataca 2986  
DB 2221 ACACGTGCTAGATTGGCGATGCAAAAGTAGCGCAACATAGAACTGGGCAATACAAATCCA 2280  
QY 2987 gacaagacaaatccaacgctgacagcatatgatataataataacacaggttttaacctta 3046  
DB 2281 GACAAAGCAAAATCCACGCTGCCAGCATTAATGATATATTAATATACAGGCTTTAACCTA 2340  
QY 3047 aaaaaataacaacccccattgactttgtctccacttatgacattgttgacttttggccaaat 3106  
DB 2341 AAAAAATAAACAACCCCATTTGCTTCCACTTATGACATTTGTGACTTTGCACTTTGCCAAT 2400  
QY 3107 ggcgaatgcacaccacgcacagtaaccatgataccgctaaacaaacacagtaaatggta 3166

DB 2401 GGCAATGCCACACCGCCACAGTAACCCATGATACCGTTAAACAAAACAGTAGTAAGTGTA 2460  
QY 3167 tatgtatgaaatgtgatatacaaccattcatctaaacggcactgatgacaataaaaaa 3226  
DB 2461 TATGATGTGAATGTGATGATACACCACTTATCATCTAACAGGCACCTGATGACAAAT 2520  
QY 3227 ctggcggtcaaaaccccaactgaacaaaacaaagtgctaatggtaatacagcaactaac 3286  
DB 2521 CTTGGCGTCAAAACCCACCAAACTGAACAAAACAAAGTGTATGTTAAATACAGCAACTAAC 2580  
QY 3287 ttttaagttaactctagtatgaagaatgcccctgtttaacgccaagaacatcgccgaataa 3346  
DB 2581 TTTAATGTTAACTCTAGTGTGAAGTGCCTTGTTAACGCCAAAAGACATCCGCCAAAAT 2640  
QY 3347 cttaaacacccctagccaagaaattcacaccaccaaagccacagcagacacccctacaa 3406  
DB 2641 CTTAAACCCCTAGCCAAAGAAATTCACACCACCAAGGACAGACACCGCCCTACAA 2700  
QY 3407 accttaccgtttaaaaggttagatgaaaataataatgtctgatgcgccaacccatcaccc 3466  
DB 2701 ACCTTTACCGTTAAAAGGTAGATGAAAATAATAATGCTGATGACGCCAAACCCATCAC 2760  
QY 3467 gtgggtcaaaagaacgcaaaataatcaagtcaaacacccctaaactcaaaaggtgaacacgt 3526  
DB 2761 GTGGGTCAAAAGAACGCAAAATAATCAAGTCAACACCCCTAACTCAAAAGGTGAAAACGT 2820  
QY 3527 cttaataataaaacccgacaaaatgttcggtttaccttggcattaaacacacagcgt 3586  
DB 2821 CTTAATATTAACCCGACAAAATGTTAGCGTTACTCTTGGCATTTAACACCAACACCGGT 2880  
QY 3587 cttaagccggcaaaagcacccctaaacacggtggtctgtctattaaaaaccccctggt 3646  
DB 2881 CTTAAAGCCGGCAAAAGCACCTTAAACGACGCTGGCTTGTCTATTAAAAACCCCACTGT 2940  
QY 3647 agcgaaacaaatccaaagtcggtgctgatggctgaagtttgcgaagtttaataatggt 3706  
DB 2941 AGCGAACAAATCCAAAGTCGGTCTGATGGCGTGAAGTTTGGCAAGGTTAATAATAATGT 3000  
QY 3707 gttgaggtgctgctgattgatggcacaactcgcattaccagagatgaattggttact 3766  
DB 3001 GTTGTAGTGTGGCATTTGATGGCACAACTCCGATTTACCAGAGATGAATTTGGCTTTACT 3060  
QY 3767 gggactaatggctcacttgataaaagcaaacccacccctaaagcaaaagcgggcaataacga 3826  
DB 3061 GGGACTAATGGCTCACTTTGATAAAAGCAAAACCCCACTTAAGCAAAAGACGGCATTAACGCA 3120  
QY 3827 ggtgtaaaagattaccaaacattcaatcaggtgagattggcccaaaacagccatgct 3886  
DB 3121 GGTGTAAAAAGATTACCAACATTCATCAGGTGAGATTGCCCAAAACAGCCATGATGCT 3180  
QY 3887 gtgacagcgcgcaagatttatgatttaaaacccgaacttgaaaaacaaatcagcagact 3946  
DB 3181 GTGACAGCGCGCAAGATTTATGATTTAAAACCCGAACCTTGAACAAACAAATTCAGCAGTACT 3240  
QY 3947 gccaaaaacagcaaaaactcattcacgaattctcagtagcagatgaacaaggttaataac 4006  
DB 3241 GCCAAAACAGCACAAAACCTCATTTACAGAAATCTCAGTAGCAGATGAACAAGGTAAATAC 3300  
QY 4007 tttacgggttagtaaccccttactccagttatgacacctcaaaagacccctctgattcatcac 4066  
DB 3301 TTTACGGTTAGTAACCCCTTACTCCAGTTATGACACCTCAAGACCCCTGATGTCATCAC 3360  
QY 4067 ttgaggtgaaaaacggcattaccaccaaggttaataaaggtggtggtggtggtggtggtggt 4126  
DB 3361 TTTGACAGGTGAAAACGGCATTTACCACCAAGGTAAATAAAGGTGTGCTGCTGTGGCATT 3420  
QY 4127 gacaaaacaaaggttcaaccacgctgaagtcgacggtgggtgggtgggttaataataatggcaagc 4186  
DB 3421 GACCAACCAAAAGGCTTAACACCGCTTAAGCTGACCGTGGGTAAATAATAATGGCAAGGC 3480  
QY 4187 attgtcattgacagcgaataatggtaaaataaccaatacagaggttaagcaaacacttagct 4246

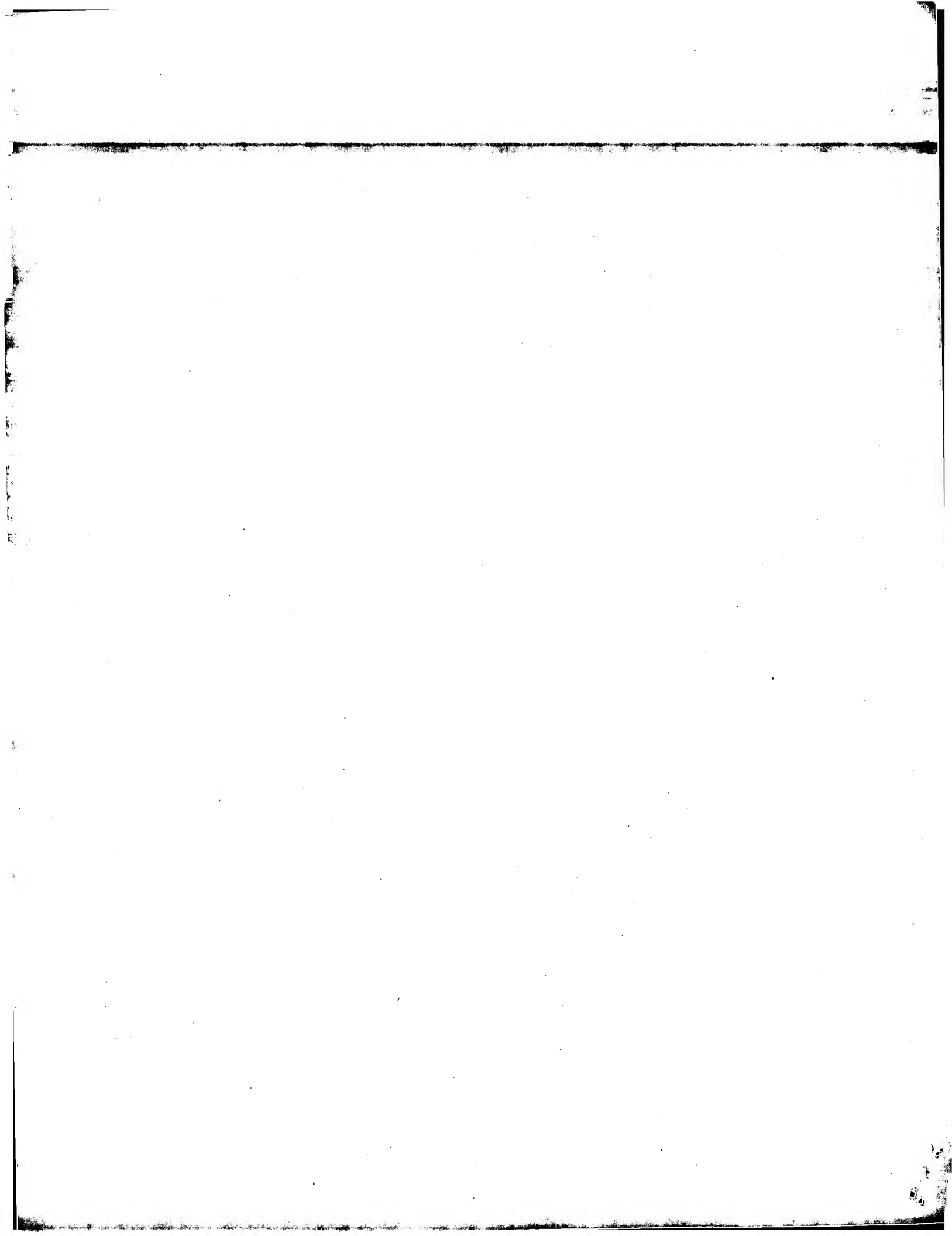


Db 3481 ATTGTCATTGACAGCCAAATGGTCAAAATACATACACAGGACTAAGCAACACTCTAGCT 3540  
Qy 4247 aatgttaccatataaaggtagcgtacgcacacacagagcaggaataataatacaaaagac 4306  
Db 3541 AATGTTACCAATGATAAAGGTAGCTGACGACACACAGAGAGGCAATATAATCAAGAC 3600  
Qy 4307 gaagacaaaacccgtgocgccagcatgttgatgtgctaaagcgcaggttttaacttgcaa 4366  
Db 3601 GAAGACAAAACCCGTGCGCGACGATTTGATGTGCTAAGCGAGGCTTTAACTTGCA 3660  
Qy 4367 gcaaatgttaacggttaacttctccacttatgacacgcgtcaactttgccgatgac 4426  
Db 3661 GGCAATGGTGAAGCGGTGACTTTGCTCCACTTATGACACCGTCAACTTTGCCGATGSC 3720  
Qy 4427 aatgccaccacccgttaaggtgacctatgatgacacaaagcaaaacccagtaagtggtctat 4486  
Db 3721 AATGCCACCCGCTAAGGTGACCTATGATGACACAAAGCAAAACACAGTAAAGTGGTCTAT 3780  
Qy 4487 gatgtcaatgtgatgatacaaccattgaagtttaagataaaaaacttggcgttaaaaaacc 4546  
Db 3781 GATGTCAATGTGGATGATACACACCTTGAAGTTAAGATTAATAAACTTGGCGTAAANACC 3840  
Qy 4547 accattgacagctactggcacaggttgctaaataaatttgcccttaagcaaatcaagctact 4606  
Db 3841 ACCACTTGCACCTACTGGCACAGGTGCTTAATAAATTTGCCCTAAGCAATCAAGCTACT 3900  
Qy 4607 ggcgatgcgcttgtaagggccagtgatcgttgctcatataaaccttatctgacgac 4666  
Db 3901 GGCGATGCGCTGTGCAAGGCCAGTGATGCTGCTCATCTTAACACCTTATCTGCGGAC 3960  
Qy 4667 atccaaactgcgaagggcgaagcgaagcgaacaaactcagcaggtctatgtggatgctgat 4726  
Db 3961 ATCCAAACTGCCAAGGGCAAGCCAAAGCAACCAACTCAGCAGGCTATGTGGATGCTGAT 4020  
Qy 4727 ggaataaggttcatctatgacagtcaccgatacaagtaactatcaagcacaataatgatgc 4786  
Db 4021 GGCAATAAGGTCACTATGACAGTACCGATACCAAGTACTATCAAGGCCAAATAATGATGGC 4080  
Qy 4787 acagttgataaaaccaaagaagtgccaaagacaaactggcgcgaagcccaaacccca 4846  
Db 4081 ACAGTTGATAAACCNAAGNAGTTGCCAAGACAACTGGTGGCCCAAGCCCAAAACCCCA 4140  
Qy 4847 gatggacattggttcgaatgaatgtcaaatcagtcattcaacaaagcaacagtaaatgat 4906  
Db 4141 GATGGCACATTGGCTCAAAATGAATGCTAAATCAGTCAATTAACAAGAACAAATGAATGAT 4200  
Qy 4907 gcaataaaaccaaagcatcaatgaagacacacgccttggtaagggacttgaaaaagcc 4966  
Db 4201 GGCAATAAAGCAAGGCATCAATCAAGACAAACGCTTTGTTAAAGGACTTGAAGAGCC 4260  
Qy 4967 gcttctgataacaaaacccaaacccgcagtaactgtgggtgatatttaattgcccgttgc 5026  
Db 4261 GCTTCTGATAAACAACCAAAACCCGCGAGTAACTGTGGGTGATTTAAATGGCGTTGCC 4320  
Qy 5027 caaacccgctgaccttgcaggggatatacagcacacgcgttaaaaaactggcgagact 5086  
Db 4321 CAAAACCCGCTGACCTTTGCAGGGGATACAGGCACAAAGGCTTAAATAAACTGGCGAGACT 4380  
Qy 5087 ttgacatacaaggtggcaacagacacacaaataagctaacccgaataataatcoggtgtg 5146  
Db 4381 TTGACCATCAAGGTTGGCAACACAGACCAATAAGTAAACCGATAATAAATCAATCGGTGTG 4440  
Qy 5147 gtagcagggtactgatggcttactgtcaaaacttgcaaaagacctaaccacttaacagc 5206  
Db 4441 GTAGCAGGTACTGATGGCTTCACTGTCAAACTTGCCAAAGACCTTACCAATTTTAAACAGC 4500  
Qy 5207 gtaatgcaggtggcaccaaaattgatgacaaagcggtgtctttttgtagactcaagcggt 5266  
Db 4501 GTTAATGCAGGTGGCACCAAAATTTGATGACAAAGGCGTGTCTTTGTAGACTCAAGCGGT 4560  
Qy 5267 caagccaaaacaaacccctctgtcaagtgcgaatggcgtggacctgggtggcaagtc 5326  
Db 4561 CAAGCCAAAGCAACACCCCTGTGCTAAGTGGCAATGGGCTGGACCTGGGTGGCAAGGTC 4620

Qy 5327 atcagtaattgtgggaaaggcacaaaagataccgaagctgccaatgtacaacagtttaaac 5386  
Db 4621 ATCAGTAAATGTGGCAAAAGGCACAAAAGATACCGACGCTGCCAATGTACAAAGTTAAAC 4680  
Qy 5387 gaagtcagcaacttgyttgggtcttgggtaattgtcgtgtaataaagcgtgacgcaaatcag 5446  
Db 4681 CAAGTACGCAACTTGTGGGTCTTTGGTAAATGCTGGTAAATGATACGGCTGACGGCAATCAG 4740  
Qy 5447 gtaaacattgcccagacatcaaaaaagacccaaatccaggttcatcatcctaaccgcactgac 5506  
Db 4741 GTAAACATTTGCCGACATCAAAAAAGACCCAAATTCAGGTTTCATCATCTAACCCGACTGTC 4800  
Qy 5507 atcaaaagcagcagcaggttacttggcggttaaggttaataaacgataccgaaaaacttgcact 5566  
Db 4801 ATCAAAAGCAGCAGCGTACTTTGGCGGTAAAGGTAAATACGATACCGAANAACCTTGGCACT 4860  
Qy 5567 ggtggtatacaagtggtggttgataaaagacgcgaacgcgtacgcgtatttaagcaatggt 5626  
Db 4861 GGTGGTATACAAAGTGGCGGTGGATAAAGACCGCAACGGTAAACGGCGATTTAAAGCAATGTT 4920  
Qy 5627 tgggtcaaaaccccaaaaagatggcagcaaaaagccctgctcgcacttaaaacgcgcga 5686  
Db 4921 TGGGTCAAAACCCAAAGATGGCAGCAAAAAGCCCTGCTCGCACCTTATAACGCCGCA 4980  
Qy 5687 ggtcagacaaactatttgacaaacccccgcagaaagcccgagagccattgacagaataaataaagca 5746  
Db 4981 GGTCAAGACCAACTATTTTGACCAACAAACCCCGCAGAACGCTTTGACNGAATAAATGMAACNA 5040  
Qy 5747 ggtatccgtcttccatgtcaacgattggcaatcaagcgctgtggtgataagggcgtaac 5806  
Db 5041 GGTATCCGCTTCTTCCATGTCAACCATGGCAATCAAGACGCTGTGTTACAAAGGGCGTAAC 5100  
Qy 5807 ggcattgactcaagtgccctcaggaagcactcagtgggagataagtttccaggaagcaagca 5866  
Db 5101 GGCATTGACTCAAGTGCCTTCAGCAAGCACTCAGTGGGGATAGGTTTTCAGGCCCAAGCA 5160  
Qy 5867 gatgtgaagccgcggttgcctatagcagacaaacccaaagcaggaacaaacccatccatcgc 5926  
Db 5161 GATGGTGAAGCCGCTTGCCTATAGCAGACAAACCCAAAGCAGGCAACCAATCCATCGCC 5220  
Qy 5927 atcgggtgataagcgaacgaacggcggtatcaatccatcgccatcgggtacaggaagatgtg 5986  
Db 5221 ATCGGTGATAAACACACAAAGCCACGGCGCATCAATCCATCCCATCGGTACAGGCAATGTG 5280  
Qy 5987 gtagcaggttaagcactctgtgcttcgcgcacccaaagcactgttaaggtctgataaagct 6046  
Db 5281 GTAGCAGGTAAAGCACCTCTGTGTGCCATCGGGACCCCAAGCAGCTGTAAAGCTGATAACAGT 5340  
Qy 6047 tacagttgggtgtaataacaaacgggtttaccgatgcccactcaaaacgatgtctttgggtg 6106  
Db 5341 TACAGTTGGGTAAATAACAAACAGTTTACCGATGCGCACTAAACCCGATGCTTTGGTGTG 5400  
Qy 6107 ggaataaactacccgtgacccgaagtaactcgttgccttaggttcaaaactctgcacac 6166  
Db 5401 GGCAATAACATCAACCGTACCGGAAAGTAACCTCGTTGCGCTTAGCTTCAAACTCTGCCATC 5460  
Qy 6167 agtgagggcacacacgcagcgcacacaaagcccaaaaaactctgacggcacagcaggtacaac 6226  
Db 5461 AGTGACGGCACACAGCGAGGCACACAAAGCCCAAAAAATCTGACGCCACAGCAGGTACAACC 5520  
Qy 6227 accacagcaggtgcaacccggtacggttaagggcttgcgtgacaaacggcggttgggtg 6286  
Db 5521 ACCACAGCAGGTGCAACCGGTACGGTTAAAGGGCTTTGCTGGACAAACGGCGGTTGGTGGC 5580  
Qy 6287 gtcctcgtgggtgctcaggttgctgaacgcgcgtatccaaaatgtggcagcaggtgaggtc 6346  
Db 5581 GTCTCCGTGGGTGCTCAGGTGCTGAACCGCGTATCCAAATGTGCCAGCAGGTGAGGTC 5640  
Qy 6347 agtgccaccagcagcaggtgoggtcaatggtagccagttgtacaaagccaccccaagcatt 6406  
Db 5641 AGTGCCACCAAGCAGCGGTGCGGTCAATGGTAGGCAGTTGTACAAAGCCACCCAAAGCATT 5700

Qy	6407	gccaacgcaaccaatgagcttgaccatcgatatccaccacaaacgaaaaataagggccaatgca	6466
Db	5701	gccaacgcaaccaatgagcttgaccatcgatatccaccacaaacgaaaaataagggccaatgca	5760
Qy	6467	gggatttcacagcgatggcgatggctccatgcccacaaagcctacattccctggcgagatcc	6526
Db	5761	gggatttcacagcgatggcgatggctccatgcccacaaagcctacattccctggcgagatcc	5820
Qy	6527	atggttaccgggggtattgcccaccacacacggtcaaggtgcggtggcgagtgaggactgtcg	6586
Db	5821	atggttaccgggggtattgcccaccacacacggtcaaggtgcggtggcgagtgaggactgtcg	5880
Qy	6587	aagctgtcgggataatggtcaatgggtatttaaaatcaatggttcagccgatacccaaggc	6646
Db	5881	aagctgtcgggataatggtcaatgggtatttaaaatcaatggttcagccgatacccaaggc	5940
Qy	6647	catgtaggggcgcgcdgttggtgcaggttttcacttt	6682
Db	5941	catgtaggggcgcgcdgttggtgcaggttttcacttt	5976

Search completed: September 13, 2001, 11:47:58  
Job time: 83299 sec





APPLICANT: Peak, Ian  
APPLICANT: Jennings, Michael  
TITLE OF INVENTION: MODIFIED SURFACE ANTIGEN  
FILE REFERENCE: 8795-24U1  
CURRENT APPLICATION NUMBER: US/09/771,382  
CURRENT FILING DATE: 2001-01-25  
PRIOR APPLICATION NUMBER: US 60/177,917  
PRIOR FILING DATE: 2000-01-25  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 19  
LENGTH: 1797  
TYPE: DNA  
ORGANISM: Neisseria meningitidis  
US-09-771-382-19

Query Match 0.8%; Score 58.6; DB 5; Length 1797;  
Best Local Similarity 51.8%; Pred. No. 0.00024;  
Matches 133; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 6429 acatcgtattcccaaaaaaaagggccaattcgagggtttcatcagcgatgcca 6488  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
DB 1541 acaaccgcacgcacaatgtggacggcaacgcgcgcgggtatgccaaagcgatgccaa 1600  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 6489 tggcgtccatgccacaagcctacattcttgcgcagatccatggttacccgggggtattgccca 6548  
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 1601 ccgcagggttgctcaggcctatttgcgcgaagagtatgatggcgatcgcgcggtta 1660  
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 6549 cccaacacgtcaaaggtgcggtgagtcagtgactgtcgaaagctgcggataaatggccaat 6608  
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 1661 ctattcgcgcgaagccgggttacgcgaatcggtactcagacattctcgacacitgggaatt 1720  
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 6609 ggggtattaaaatcaatggttcagccgatcccaagcccatgtagggggcgaggtgggtg 6668  
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 1721 gggttatcaaggcacggcttcgcgaattcgcgcgggttacttcggtacctccgcacatcg 1780  
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 6669 cagggttttcacttttaa 6685  
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 1781 tcggttatcagtggttaa 1797  
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3  
US-09-771-382-22  
Sequence 22, Application US/09771382  
GENERAL INFORMATION:  
APPLICANT: Peak, Ian  
APPLICANT: Jennings, Michael  
TITLE OF INVENTION: MODIFIED SURFACE ANTIGEN  
FILE REFERENCE: 8795-24U1  
CURRENT APPLICATION NUMBER: US/09/771,382  
CURRENT FILING DATE: 2001-01-25  
PRIOR APPLICATION NUMBER: US 60/177,917  
PRIOR FILING DATE: 2000-01-25  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 22  
LENGTH: 1815  
TYPE: DNA  
ORGANISM: Neisseria meningitidis  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: "n" is any nucleotide, synonymous nucleotide or absent nucleotide  
OTHER INFORMATION: at a corresponding position in any one of SEQ ID NOS:12-22  
US-09-771-382-22

```

Db 1559 acaaccnctcgcacaaatgtgnacgagcaacgcgcgcnagcgggnatcgcgcccaagcagattgcaa 1618
QY 6489 tggcgtccatgccacaagcctacattctctggcagatccatggttacccgggggtattgcca 6548
Db 1619 ccgcaggtntngntcagcgcgtatntgcccgcgaagagatgatggcgatcgcgcgcnna 1678
QY 6549 cccacaacggttcaaggctgcgggtggcagtgggagactgctcgaagcgtctcggataaatgggtcaat 6608
Db 1679 ctatcncgcgcgaagcgggttangccatcgctactcgnagnattcngnncnngnnaatt 1738
QY 6609 ggggtatttaaatcaatgggttcagccgatacccaagggcctagttagggcgcgagttgggtg 6668
Db 1739 ggnntatcaangcacggcgttccgcgaattcgcgcgncatttcggttcttcgcgcatctg 1798
QY 6669 cagggttttcaacttttaa 6685
Db 1799 tcggttatcantggttaa 1815

RESULT 4
US-09-771-382-30
; Sequence 30, Application US/09771382
; GENERAL INFORMATION:
; APPLICANT: Peak, Ian
; APPLICANT: Jennings, Michael
; TITLE OF INVENTION: MODIFIED SURFACE ANTIGEN
; FILE REFERENCE: 8795-24U1
; CURRENT APPLICATION NUMBER: US/09/771,382
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 60/177,917
; PRIOR FILING DATE: 2000-01-25
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 1224
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-09-771-382-30

Query Match 0.8%; Score 57; DB 5; Length 1224;
Best Local Similarity 51.4%; Pred. No. 0.0005;
Matches 132; Conservative 0; Mismatches 125; Indels 0; Gaps 0

QY 6429 accatcgatccaccacaaacgaaaaaaagggccaaatgcagggattttcatcagcgatggcga 6488
Db 968 acaacgcgctcgacaatgtggacggcaacgcgctgcgggcacgcgccaaagcgtattgcaa 1027
QY 6489 tggcgtccatgcccacaagcctacattccttgcgagatccatggttaccgggggtattgcca 6548
Db 1028 ccgcaggtcgtgttcaggcgtattttccgcgaagagtatgattggcgatcgcgcgcggcca 1087
QY 6549 cccacaacggtcaagtgcggtgctgcggtggagctgcgaagctgtcggataatgggtcaat 6608
Db 1088 ctatcgcgcggaagccggttaacgccatcggctacgccagattttccgcgcggcgaaatt 1147
QY 6609 ggggtatttaaatcaatggttcgcgcgatacccaagggcctagttagggcgcgagttgggtg 6668
Db 1148 ggattatcaaggaacggcgttcgcggaattcgcgcggccatttcggtgtcttcgcgcatctg 1207
QY 6669 cagggttttcaacttttaa 6685
Db 1208 tcggttatcantggttaa 1224

RESULT 5
US-09-771-382-31
; Sequence 31, Application US/09771382
; GENERAL INFORMATION:
; APPLICANT: Peak, Ian
; APPLICANT: Jennings, Michael
; TITLE OF INVENTION: MODIFIED SURFACE ANTIGEN
; FILE REFERENCE: 8795-24U1

```

## 7. RESULT

```

US-09-771-382-28
; Sequence 28, Application US/09771382
;
; GENERAL INFORMATION:
;
; APPLICANT: Peak, Ian
;
; APPLICANT: Jennings, Michael
;
; TITLE OF INVENTION: MODIFIED SURFACE ANTIGEN
;
; FILE REFERENCE: 8795-24U1
;
; CURRENT APPLICATION NUMBER: US/09/771,382
;
; CURRENT FILING DATE: 2001-01-25
;
; PRIOR APPLICATION NUMBER: US 60/177,917
;
; PRIOR FILING DATE: 2000-01-25
;
; NUMBER OF SEQ ID NOS: 52
;
; SOFTWARE: PatentIn version 3.0
;
; SEQ ID NO 28
;
; LENGTH: 1539
;
; TYPE: DNA
;
; ORGANISM: Neisseria meningitidis
;
US-09-771-382-28

```

US-09-771-382-32  
; Sequence 32, Application US/09771382

Db	1343	ccgcaggctcgttctcaggcgctattgcccgcaagagtatgatggcagtcgcgcgcgca	1402
Qy	6549	cccacaacggtcaagctgcggctggcagctgtcgaaagctgtcgataatggtccaat	6608
Db	1403	cttatcgcgcggaagcggcttacgcgcacatcggtactcagtatcttcgacgcggaaatt	1462
Qy	6609	gggtattataaatacaatggttcagccgcataccaaggccatctaggggcgcgagttgggt	6668
Db	1463	ggattataaaggcagcggttcgcggcaattcgcggccatttcggtgtcttcgcgcatctg	1522
Qy	6669	caggttttcaacttttaa	6685
Db	1523	tcggttatacagtggttaa	1539

US-09-771-382-14  
; Sequence 14, Application US/09771382

```

? APPLICANT: Jennings, Michael
? TITLE OF INVENTION: MODIFIED SURFACE ANTIGEN
? FILE REFERENCE: 8795-24U1
? CURRENT APPLICATION NUMBER: US/09/771,382
? CURRENT FILING DATE: 2001-01-25
? PRIOR APPLICATION NUMBER: US 60/177,917
? PRIOR FILING DATE: 2000-01-25
? NUMBER OF SEQ ID NOS: 52
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 14
? LENGTH: 1770

```

```

Query Match          0.8%; Score 57; DB 5; Length 1779;
Best Local Similarity 51.4%; Pred. No. 0.0006;
Matches 132; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```





US-09-11-382-29

bb 1526 tcggttatcagtggttaa 1542

Search completed: September 13, 2001, 12:35:56  
Job time: 84918 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2001, 13:49:24 ; Search time 78.73 Seconds  
(without alignments)  
1576.239 Million cell updates/sec

Title: US-09-361-619-7  
Perfect score: 10303  
Sequence: 1 MNHYKVFNRATGTFMAVA.....NGSADTQCHVCAAVGAGPHF 2047

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0601.\*  
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10303	100.0	2047	22	AA669134 M. catarrhalis str
2	10024	97.3	1992	22	AA669133 M. catarrhalis str
3	10021	97.3	1992	17	AAW04505 Moraxella 200 kDa
4	10021	97.3	1992	22	AA669137 M. catarrhalis M56
5	7675	74.5	2033	22	AA669135 M. catarrhalis str
6	6593.5	64.0	2314	22	AA669136 M. catarrhalis les
7	1123	10.9	2353	17	AA693933 Haemophilus adhesi
8	1109	10.8	2411	21	AA623860 Haemophilus influe
9	1016	9.9	1104	21	AA623856 Haemophilus influe
10	1016	9.9	1104	21	AA623859 Haemophilus influe
11	970	9.4	1004	21	AA623857 Haemophilus influe

12	927.5	9.0	1002	21	AA623854 Haemophilus influe
13	813	7.9	2042	19	AAW56319 Haemophilus paraga
14	804	7.8	2039	19	AAW56322 Haemophilus influe
15	697.5	6.8	1094	21	AA623858 Haemophilus adhesi
16	678.5	6.6	1098	17	AA693932 Haemophilus adhesi
17	573	5.6	679	17	AA693934 Haemophilus influe
18	573	5.6	679	21	AA623855 Haemophilus influe
19	520.5	5.1	1601	18	AAW30292 Non-typeable Haemo
20	505.5	4.9	1598	18	AAW30291 Non-typeable Haemo
21	496.5	4.8	1529	14	AA641732 High molecular wei
22	478.5	4.6	2383	21	AA615945 E. coli proliferat
23	477.5	4.6	2514	21	AA675097 Neisseria meningit
24	462	4.5	2893	19	AAW98828 H. pylori GPHO 148
25	462	4.5	2893	19	AAW98828 Helicobacter polyp
26	460.5	4.5	1978	20	AAW27230 Amino acid sequenc
27	457	4.4	1338	14	AA641731 High molecular wei
28	452	4.4	2599	21	AA675098 Neisseria meningit
29	451	4.4	3647	11	AA605041 Filamentous haemag
30	450.5	4.4	1638	20	AAW00138 Enterococcus faeca
31	450.5	4.4	1638	20	AAW00140 Enterococcus faeca
32	450.5	4.4	1638	20	AAW00142 Enterococcus faeca
33	448	4.3	3596	21	AAW87407 Bordetella pertuss
34	443.5	4.3	1962	12	AA610563 Mutant protease (K
35	443	4.3	1221	21	AA601825 Haemophilus influe
36	443	4.3	1227	21	AA601824 Haemophilus influe
37	439.5	4.3	1962	12	AA610560 Mutant protease (K
38	439.5	4.3	1968	12	AA610941 Mutant protease (d
39	438.5	4.3	1962	12	AA610561 Mutant protease (N
40	438.5	4.3	1981	19	AAW42634 Protein sequence t
41	438	4.3	1959	12	AA610562 Mutant protease (d
42	437.5	4.2	1962	12	AA610557 Mutant protease (A
43	436.5	4.2	1962	12	AA610558 Mutant protease (A
44	436.5	4.2	1962	12	AA610559 Mutant protease (A
45	435	4.2	1612	19	AAW65088 R. prowazekii S-la

## ALIGNMENTS

RESULT 1  
AA669134  
ID AA669134 standard; Protein; 2047 AA.  
XX  
AC AA669134;  
XX  
XX 24-APR-2001 (first entry)  
XX  
XX M. catarrhalis strain 4223 genomic 200kda protein SEQ ID NO:7.  
XX  
XX Moraxella catarrhalis strain 4223; major outer membrane protein;  
KW 200kda outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection.  
XX  
XX Moraxella catarrhalis.  
XX  
XX WO200107619-Al.  
XX  
XX 01-FEB-2001.  
XX  
XX 26-JUL-2000; 2000WO-CA00870.  
XX  
XX 27-JUL-1999; 99US-0361619.  
XX  
XX (CONN-) CONNAUGHT LAB LTD.  
XX  
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;  
XX WPI; 2001-159722/16.  
XX N-PSDB; AAF59102, AAF59103.  
XX  
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
XX useful in protective vaccines and for diagnosis -

Claim 1; Fig 3A-W; 247pp; English.

PS The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. CC The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II), and its truncated versions, are used as CC immunogenic compositions and vaccines to protect against M. catarrhalis CC infections, particularly otitis media in humans. (II) is also used as CC antigen in immunoassays for detecting specific antibodies (Ab), and to CC generate Ab. (I) are used for recombinant production of (II) and its CC fragments are used as probes for identifying/cloning 200 kDa protein CC genes from other strains, and for diagnostic detection of M. catarrhalis. CC (I) makes possible production of large amount of recombinant immunogens. CC Expression of truncated versions of (II) reduces toxicity of the protein CC towards the Escherichia coli host. The present sequence represents the CC M. catarrhalis strain 4223 genomic 200kDa protein, which is given in the CC exemplification of the present invention.

XX Sequence 2047 AA;

Query Match 100.0%; Score 10303; DB 22; Length 2047;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2047; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNHLYKVFENKATGTFMAVEAKSHSTGGSCATQGVSVCTLSFARLAALVIVIGAT 60  
DB 1 mnhlykvfena cgtfmaveakshstggscatqgvsvctlsfariaalavlivigat 60  
QY 61 LSGSAYAQKDKTHIAIGBQNPRRSGTAKADGDRAIAGENAGGQAIAGSSNKTV 120  
DB 1 lsgsayaqkdkthiaigebnprrsgtakadgdraiaigenanaggqaiagssnktv 120  
QY 121 NGSSLDKIGTDATQBSIAIGDVKASGDASTAIGSDDLHLDDQHGNPKHPKGTTLNDLI 180  
DB 121 ngssldkigtatqbsiaigdvkasgdastaisgddhlldqhgnpkhpkgttlndli 180  
QY 181 NGHAVLKEIRSSKNDPVKYRRYRTTASGHASTAVGAMSYAOGHFSNAPGTRATAKSAYSLAV 240  
DB 181 nghavlikeirsskndpvkyrrytasghastavgamasyaoghfsnagtrataksayslav 240  
QY 241 GLAATAGGOSTIAIGSDATSSSLGATLACAGTRAOIQQSIALCOGSVVVTQSDNNRSPAYT 300  
DB 241 glaataeggstiaigsdatssslgaialagagtraoigqsiacogsvvvtqsdnnrspayt 300  
QY 301 PNTQALDPKFOATNPKAGPLSTGSSNSIKRKIINVAGYKNTDANVAQLEAVVWAKER 360  
DB 301 pntqaldpkfoatnkpagplstgssnsikrkiinvagynktadanvaqleavvwaker 360  
QY 361 RITFQGDNDSTVYKIGLDNLTIKGAEETNALTDNNIGVVKADNSGLKVKLAKTLNNLT 420  
DB 361 ritfggdndstvykigldnltikgaetnaltdnnigvkvadnsglkvklaktlnnlt 420  
QY 421 EVNTTTLNATTVYKVGSSSTTAELLSDSLTTPQNTGSGQSTKTYGVNGYKFTNNAET 480  
DB 421 evntttlntatvtykvgssttaellsdslttpqntgsgqstktvgvngykftnnat 480  
QY 481 TAAIGTRITTRDKIGFARDGDVDEKQAPYLDKQKLVKGSVAITIDNGDAGNKKISNLAK 540  
DB 481 taagtritrtdkigfardgdvdekqapyldkkqlkvgsvaitidngdagnkksnlak 540  
QY 541 GSSANDAVTIEQLKAAKPTLNAGAGTSVPTTETSVDKASGNVTAPTNYIGVKTTELSNDG 600  
DB 541 gssandavtieqlkaakptlnagagtsvpttetsvdkasgnvtaptnyigvkttelnsdg 600  
QY 601 TSDKFSVKGSGTNNLSLVAEHLASLYNEVNRTPADSALQSFTVKEEDDDANAITVAKDPT 660  
DB 601 tsdkfsvkgsgtnnlsylvaehlasynevnrtpadsalqsfvkeedddanaitvakdpt 660  
QY 661 KNAGAVSILKLGKNGLTVAATKDGTVTFGLSQDSGLATIGKSTLNDGLTYKDTNEQIQV 720  
DB 661 knagavsilklgkngltvatkdgdtvtfglsqdsglatigkstlndgltykdtneqiqv 720

QY 721 GANGIKFTVNGSNPCTGTANTARITRDKIGFAGSDGAVDTNKNPYLDQDKLQVGNVKITN 780  
DB 721 gangikftvngsnpgtgiantaritrtdkigfagsdgavdtnkpyldqdklqvgvgnvkitn 780  
QY 781 TGINAGSKAITGLSPPLPSIADQSSRNIEGNTIQQDKSNAASINDILNTGFENLNNNN 840  
DB 781 tginagskaitglspplsiaadqssrniegtiqgdkksnaasindiintgfnlknann 840  
QY 841 PIDFVSTYDIVDFANGNATTATVTHDTANKTSKVVDVNVDDTTIHLTGTDDNKKLGVT 900  
DB 841 pidfvstydivdfangnattatvthdtanktskvvydvnvddttihltgtdnnkklgvt 900  
QY 901 TKLNKTSANGNTATNPNVNSDEDAVLNNAKDAENLTLAKEIHTHTKGPADTALQFTVYK 960  
DB 901 tklnktsangntatnfnvnsdedalvnakdaenlntlakeihthtkgpadtalqftvk 960  
QY 961 KYDENNNADANAITVGOKNANNQVNTLTKGENGLINIKTDKNGTVTFGINTTSLGAKG 1020  
DB 961 kydennnadana itvgknannqvntltkgennglniktdkngtvtfginttsglakg 1020  
QY 1021 STLNDGGLSIKNPTGSEQIQVGADGVKFAKVNNGVVGAGIDGTTRITRDEIGFTGTNGS 1080  
DB 1021 stlndgglisiknptgseqiqvgadvkfakvnnngvvgagidgttritrdeigftgtngs 1080  
QY 1081 LDKSPHLSKDGINAGCKKITNIQSGEIAQNSHDVATGCKIYDLKTELENKTSSTAKTAQ 1140  
DB 1081 ldksphliskdginagckkitniqsgeliaqnshdavatgckiydlktelenktsstaktaq 1140  
QY 1141 NSLHFFSVADEOGNFTVSNPSSYDTSKTSDVITFAGENGITTKVNKGVRVGIDOTKG 1200  
DB 1141 nslhffsva deognftvsnpsysdytsktsdvitfagengittkvnkgvrvgidotkg 1200  
QY 1201 LTPKLTGVNNGKGVIVDSQNGQNTITGLSNTLANVTNDKGSVRTTEOGNIKDEKTR 1260  
DB 1201 ltpkltgvnngkgv ivdsqngqntitglsntlanvtn dkgsvrtteognikdedktr 1260  
QY 1261 AASIVDVLISAGFNLOGCEAVDFVSTYTFVNFADGNATTAKVTYDDTSKTSKVVDVNV 1320  
DB 1261 aasivdvlisag fnlognceavdfvstytfvnf adgnattakvtyddtsktskvvydvnv 1320  
QY 1321 DTTIEVKDKLGVKTTTLTSTGTGANKFALSNOATGDLAKASDIVAHLNLTSGDITQAK 1380  
DB 1321 dttievkdklgvkttltstgtgankfalsnogatgdlavkasdivahlntlsgdiqtak 1380  
QY 1381 GASQANNAGYVDADGNKVIVYSTDNKYQAKNDGTVDKTEKVAKDKLVAQATPDGTLA 1440  
DB 1381 gasqannag yvdadgnkv ivystdnkyqakndgtvdktekvakdklvagatpdgtla 1440  
QY 1441 OMNVKSVINKEOVNDANKKOGINEDNAPVKGLEKAASDNKTKNAATVYGDNLNAVAQPLT 1500  
DB 1441 omnvksvinkeov ndankkoginednapvkglekaasdnktnaatv ygdnlnavaqplt 1500  
QY 1501 FAGDTGTTAKLGETLTIKGQTDNTNKLTDNNIGVYVAGTGTGFTVKLAKDLTLNLSVNA 1560  
DB 1501 fagdtgttaklgetltikgqtdntnkltdnnigvvyvagtgtgftvklakdltnlsvnag 1560  
QY 1561 TKIDKGVSVFVSDSSCOAKANTPVLISANGLDGCKVIVSNVKGKTKDTDAANVOQLNEVR 1620  
DB 1561 tkidkgvsvfv sdscoakantpvlisangldgckv ivsnvkgktktdaanvoqlnevr 1620  
QY 1621 LGLGNAGNDNADGNNOVNIADTKDPNSGSSSRNRTVIKAGTVLGGKNNDTKLTATGGIQV 1680  
DB 1621 lglgnagn dndnagdnovniadt kdpnsgsssrnrtvikagtv lggknnntekltatggioq 1680  
QY 1681 GVDKGNANGDLSNVVKTQDKSKKALLATYNAAGQNTYLTNNPAAEIDRINEQIRFF 1740  
DB 1681 gvdkgdnang dlsnvvktqdkskallaty naagqntyltnnpaaeidrineqirff 1740  
QY 1741 HVNDGNQPVVQGRNGIDSSASGKHSVAIGFOAKADGAAVAIGRQTQAGNQSIAIGNA 1800  
DB 1741 hvndgnqp vvqgrngidssasgkhsvaigfoakadgaavaig rqtqagnqsiaign 1800  
QY 1801 QATGDSIAIGTVNVVACKHSGAIGDPSTVKADNSY SVGNNNQFTDTATQTDVFGVGNIT 1860



QY 1076 GTNGSLDKSPHLSKOGINAGGKKTINIOGGEIAQNSHDAVTCGKIYDLEKISST 1135  
Db 1021 gtngsldkskphlskdglnagggkktiniqgeiaqnsdhavtggkiydktelenkisst 1080  
QY 1136 AKTAQNSLHEFSVADRGNNFTVSNPYSSVDTSTKSDVTFAGENGITTKVKNKGVVVRVGI 1195  
Db 1081 akcaglnslhefsvadeggnftvsnpyssvdtstksdvtfagengitckvknkgvvrvgi 1140  
QY 1196 DQTKGLTTPKLTGNNNGKGIIVDSQNGQNTITGLSNTLANVTNDRKGSVRTTEQGNIIKD 1255  
Db 1141 dqtkgltpkltvgnnngkgiivdsngqntitglsntlanvtnndkgsrvrtteqgnlikd 1200  
QY 1256 EDKTRAASIVDLVSAGFNLOGNEAVDFVSTYDTPNADGNATTAKVTYDDTSKTSKVY 1315  
Db 1201 edktraasivdlvsagfnlqngneavdfvstydtpnfdagnattakvtyddtsktskvvy 1260  
QY 1316 DVNVDDTTIRVDKDKLGVTTLTSTGTCANKPALSNOATGDAVKASDITVAHLNLTSGD 1375  
Db 1261 dvnvddttirvdkdklgvttltstgtcankpalsnogatgdavkassdltvahlntlskd 1320  
QY 1376 IQTAKASQANNSAGYVDADGNKVIYDSDNKYIQAKNDGTVDKTKEVAKDKLVAQAQTP 1435  
Db 1321 iqtakasqannsagvydadgnkviydstdnkyyqakndgtvdktkevakdklvaqaqtp 1380  
QY 1436 DGTLAQNNVKSVINKEQVNDANKQGINEDNAPVKGLEAASDNKTKNAAVTVGDLNAVA 1495  
Db 1381 dgtlaqnnvksvinkeqvndankqginednapvkgleaasdnktknaavtvgdlnava 1440  
QY 1496 QTPLTTFAGDTGTTAKLIGFTLTITKGGOTDTNKLTDNNIGVVGAGTGTGTVKLAKDLNLNS 1555  
Db 1441 qtplttfagdtgttacklignftltitkggottcnkltddnigvvgagtggtvklakdlnlns 1500  
QY 1556 VNAGGKTIDDKGVSFVDSGQAKANTPVLSANGLDLGGKVISNVGKTGKTDAAVYQQLN 1615  
Db 1501 vnaggktiddkgvsvfvsdgqakantpvlsangldlggkvisnvvgktgktdaavvqqln 1560  
QY 1616 EVRNLLGLGNAGNDADGNQVNIADTKDPNSGSSNRVIRAGTVLGGKGNNDTEKLAT 1675  
Db 1561 evrnllglnagnadnagnvniadtkdpnsgssnrvtvikagtvlggkgnndteklat 1620  
QY 1676 GGTOGVGDGNGANGDLSNVWKTOKDGSKALLATYNAAAGOTNYLTNNPAEAIDRINEQ 1735  
Db 1621 ggtogvgdgngangdlsnvwktokdgskallatynaaagotnyltnnpaeaidrineq 1680  
QY 1736 GIRFFHVNDGNQBPVYVGRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNQSTA 1795  
Db 1681 girffhvnngnqbpvyvgrngidssasgkhsvaigfoakadgeaavaigrqtqagnqsta 1740  
QY 1796 IGDNAQATGDSIAIGTGNVVGKHSAGTCDPSTVKADNSYSVGNNGQTTDATQDVFV 1855  
Db 1741 igdnaqatgdsiaigtgnvvvgkhsagtdpstvkadnsysvgnngqttatqdvfv 1800  
QY 1856 GNNITVTESNVALGSNSAITSAGTHAGTQAKKSDGTAGTTTATAGTGVKGFAGQTPAVGA 1915  
Db 1801 gnnitvtesnvalgsnsaisethagtkaksdgtagttttagtcgvkqfagqtpavga 1860  
QY 1916 VSVGASGAERRIQVWAAGEYSATSTDAVNGSOLYKATQSIANATNLDHRIHQENKANA 1975  
Db 1861 vsvgasgaerriqvwaagesatstdavngsqlykatqsianatneldhrihqenkana 1920  
QY 1976 GISSAMAMASMPQAYIPGRSMVTGGTATHNGQCAVAVGLSKLSDNGQWVFKINGSDTQ 2035  
Db 1921 gissamamasmpqayipgrsmvttggtathngqgavavglsklsgnqgwfkingssdtq 1980  
QY 2036 HVGAAYVAGGFHF 2047  
Db 1981 hvgaayvaggfhf 1992

RESULT 3  
AAW04505  
ID AAW04505 standard; Protein; 1992 AA.

XX AAW04505;  
AC 25-JAN-1997 (first entry)  
XX Moraxella 200 kDa outer membrane protein.  
DE Outer membrane protein; OMP; immunogen; vaccine; otitis media;  
XX diagnosis.  
KW Mycobacterium catarrhalis strain 4223.  
KW OS  
XX WO9634960-A1.  
XX PN  
XX 07-NOV-1996.  
XX PD  
XX 29-APR-1996; 96WO-CA00264.  
XX PF  
XX 26-MAR-1996; 96US-0621944.  
PR 01-MAY-1995; 95US-0431718.  
PR 07-JUN-1995; 95US-0478370.  
XX PA  
XX (CONN-) CONNAUGHT LAB LTD.  
XX PI  
XX Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;  
XX WPI; 1996-506162/50.  
XX DR  
XX N-PSDB; AAT38740.  
XX Moraxella outer membrane protein - useful as immunogen in protective  
XX vaccine and for diagnosis  
XX Claim 14; Fig 6; 109pp; English.  
XX An approx. 200 kDa outer membrane protein (AAW04505) can be  
XX isolated from Moraxella catarrhalis otitis media strain 4223  
XX by electroelution, or expressed from a gene (see also AAT38740)  
XX outer membrane protein is useful as an immunogen to protect  
XX against infection by Moraxella, esp. M. catarrhalis. It can  
XX also be used to detect antibodies, esp. for differential diagnosis  
XX between bacteria that cause similar symptoms, and also useful as  
XX a carrier for other antigens and used to raise antitumour  
XX antibodies for conjugation to therapeutic agents.  
XX Sequence 1992 AA;  
Query Match 97.3%; Score 10021; DB 17; Length 1992;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1991; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 56 VIGATLSSGAYAQKDKTHIAIGEQQNRRSGTAKADGDRATIAIGENANAQGGQATIGS 115  
Db 1 mlgatlgssayaqkdkthiaigeqqnrrsgtakadgdratraigenaanaqggqatig 60  
QY 116 SNKTVNGSLDKIGTDTATGOSTAIGDVKASGDASIAIGSDDLHLLDQHGNPKPKGTL 175  
Db 61 snktvngssldkigtatgostaiigdvkasgdasialgsddhlldqghnphkpgtl 120  
QY 176 INDLNGHVLKLEIRSSKNDVKYRPTTASGHASTAVGAMSYAQGHFSNAFCTRATAKSA 235  
Db 121 indlngthavleirsskndvkyrrttasghastavgamsyaqghfsnafctrataka 180  
QY 236 YSLAVGLAATAEGQSTIAIGSDATSSSLGAIALGAGTQALQGLSIALGSGSVVTTQSDNNS 295  
Db 181 yslavglataegqstiaigsdatssslgaialgagtrqqlgsgsvvtqsdnns 240  
QY 296 RPAYTPNTQALDPKFOATNTKAGPLISIGNSSTIKRKIIINVAGVKNKTDAVNAQLEAVYK 355  
Db 241 rpaytpntqaldpkfoatntkagplisignsstikrkiiinvagvknktdavnaqlavvk 300  
QY 356 WAKERRITFGDDNSTVDKIGLNDTLITIKGGAETNALTNNIGVVKKEADNSGLKVKLAKT 415

Db 301 wakeriffggdnstvdvkgldntltikggaecnaldnngivvkeadnsglkvklakt 360  
Qy 416 LNNTEVNTTFLNATTVTKVSSSTTAELLSDSLTFTQPNVTGSGSTKTYGVNGVKFT 475  
Db 361 lnnitevnttlnattvkvsgsssttaellsdsltftqpnvtgsgstektvgyngvkft 420  
Qy 476 NNAETTAAGTTRITROKIGFARGDGVDEKOAPYLDKKQLKGVSAITDNGIDAGNKKI 535  
Db 421 nnaetaaigttrtdkigfardgvddekqpyldkkqlkvgsvaltidngidagnkkl 480  
Qy 536 SNLAKGSSANDAVTEQLAAKPTILNAGAGISVTPTELSVDKAGSGNVTAPYNTIGVKTT 595  
Db 481 snlakgssandavteqlaaakptilnagagisvtppteelsvdaksgnvtapynigvktt 540  
Qy 596 LNSDGTSDKFSVKGSGTNSLVTAEHLASYLENEVNRITADSLAQSFYKEDDDDDANATTV 655  
Db 541 lnsdgtSDKFSVKGSGTNSLVTAEHLASYLENEVNRITADSLAQSFYKEDDDDDANATTV 600  
Qy 656 AKDTTKNAGAVSILKLGKNGLTVAATKKGTVTFGLSODSGLTIGCKSLNNDGLTVKDTN 715  
Db 601 akdttknagavsilklgkngltvatkkgdgtvtfglsqdsdgtlgtkstinndgltkvtdn 660  
Qy 716 EGIQVANGIKFTVWNGSNPCGTANTARITPRDKIGFAGSDGAVDTNKPYLDDQDKLQVGN 775  
Db 661 egivgngikftvwnsgnpgtgtantaritprdkigfagsdgavdtnkpyldqdklqvgv 720  
Qy 776 VKIYNTGINAGKAITGLSPRLPSTADQSSRNIELGNTIQDKDSNAASINDILNTGNL 835  
Db 721 vkityntginagkaitglspplsdiadqssrniegntiqdkdsnaasindiintgfnl 780  
Qy 836 KNNNPIDFVSTYDVIDPANGNATVAVTHDTANKTSKVYDVNVDDFTIHLTGFDNNK 895  
Db 781 knnnpidfvtstydyidpangnatvatvthdtanktskvdyvndvddftihltgtdnnk 840  
Qy 896 LGVTKTLNKTSAACNTATNFVNSDDEALVNAKDAENLTLAKEIHTTKGTADTALQ 955  
Db 841 lgvtktklnktsangntatnfvnssddealvnakdtaenltlakeihttkgtadtalq 900  
Qy 956 TFTVKVDENNADANAITYGOKNANNQVNTLTILKGSEGLNITDKRNGTVTFGINTTSG 1015  
Db 901 tftvkvdennadanaityvgoknannqvnlttikgseglntdkrngtvtfginttsg 960  
Qy 1016 LKAKSTLNDGGLSTKNPTGSEQIQVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFT 1075  
Db 961 lkakstlndgglstknptgseqiqvgadgvkfakvnngvvgagidgttritrdeigft 1020  
Qy 1076 CTNGSLDKSKPHLSKDGINAGKKITNIQSGEIAQNSHDAVTGGKIYDLKTELENKISST 1135  
Db 1021 gtngslDKSKPHLSKDGinagkkitniqsgelagqnsdhavtggkiydlktelenkisst 1080  
Qy 1136 AKTAQNSLREFSVADEQGNFTVSNPYSSTYDTSKTSQVITFAGENGITTKVNGKGVVRVGI 1195  
Db 1081 aktagnslhefsvadeqgnftvsnpysstysdtsktsqvitfagengittkvngkvrvvgi 1140  
Qy 1196 DOTKGLTTPKLTIVGNNGKGVIDVSONQNTITGLSNLTANVTNDKGSVRTTEQGNIIKD 1255  
Db 1141 dqtkglttPKLTIVGNNGKGVIDVSONQNTITGLSNLTANVTNDKGSVRTTEQGNIIKD 1200  
Qy 1256 EDKTRAASIVDVLSAGFNLONGEAVDFVSTYDVTNFPADGNATTAKVYDDTSKTSKVY 1315  
Db 1201 edktraasivdvlsagfnlNGEAVDFVSTYDVTNFPADGNATTAKVYDDTSKTSKVY 1260  
Qy 1316 DVNYDDTTIEVKDKKLGKVTTLTSTGTGANKFALSNGATGDALVKASDIVAHLNLTLSGD 1375  
Db 1261 dvnyddttievdkkkgkvtltstgtgankfalsngatgdalvkasdiVAHLNLTLSGD 1320  
Qy 1376 TOTAKGASQANNSAGYVDADGNKVYIDSTDNKYQAKNDGTVDKTEKAVAKDLVAQAQTP 1435  
Db 1321 totakgasqannsagYVDADGNKVYIDSTDNKYQAKNDGTVDKTEKAVAKDLVAQAQTP 1380  
Qy 1436 DGTLAQMNVKSVINKEQVNDANKQGINEDNAFVKGLEKASDNKTKNAAVTVGDLNAYA 1495  
|||||

Db 1391 dgtlaqmnmvksvinkeqVNDANKQGINEDNAFVKGLEKASDNKTKNAAVTVGDLNAYA 1440  
Qy 1496 QTPLTFAAGDTGTAKKLGELTITIKGGQTDITNKLTDNNIGVVAGTDGFTVKLAKDLTNLNS 1555  
Db 1441 qtpltfaagdtgtakklgetilitikggqtditnkltdnnigvvagtdgftvklakdltnlns 1500  
Qy 1556 VNAGGTTKIDDKGVSPVSDSSGQAKANTPVLSSANGLDLGGKVISNVKGTKYDADAANVQQLN 1615  
Db 1501 vnaggtkidDKGVSPVSDSSGQAKANTPVLSSANGLDLGGKVISNVKGTKYDADAANVQQLN 1560  
Qy 1616 EVRNLLGAGNAGNDNADGNQVNIADIKKDPNSGSSNRTVIKAGTVLGKGNNDTEKLTAT 1675  
Db 1561 evrnllgagndnagndgnqvnIADIKKDPNSGSSNRTVIKAGTVLGKGNNDTEKLTAT 1620  
Qy 1676 GGIOGVGDKDNANGDLNVVWVKTKQDGSKKALLATYNAAGQNTYLTNNPAEADIRINEQ 1735  
Db 1621 ggiovgvdkdngnangdlNVVWVKTKQDGSKKALLATYNAAGQNTYLTNNPAEADIRINEQ 1680  
Qy 1736 GIRFFHVNDGNOEPVQQRNGIDSSASGKSHVAIGFQAKADGEAAVAIGRQTQAGNQSTIA 1795  
Db 1681 girffhvndgnoepvqqrngidssasgkshvalgFQAKADGEAAVAIGRQTQAGNQSTIA 1740  
Qy 1796 IGDNAQATGDDOSIATGTGNVAGKHSAGTGPSTVKADNSYSVGNNOFTDATOTDVRGV 1855  
Db 1741 igdnaqatGDDOSIATGTGNVAGKHSAGTGPSTVKADNSYSVGNNOFTDATOTDVRGV 1800  
Qy 1856 GNNITVTESNSVALGSNSAISAGTHAGTQAKKSDGTAGTATTTAGATGTVKGFAGTAVGA 1915  
Db 1801 gnnitvtesnsvalgsnsaisagthagTQAKKSDGTAGTATTTAGATGTVKGFAGTAVGA 1860  
Qy 1916 VSVGASGAEERRIQNVAAGEVSATSDAVNGSOLYKATQSIANATNELDHRITHQENKANA 1975  
Db 1861 vsvgasgaeRRIQNVAAGEVSATSDAVNGSOLYKATQSIANATNELDHRITHQENKANA 1920  
Qy 1976 GISSAWANASMPQAVIPGRSMVTGCIATHNGGAVAGLSKLSNDGQVFKINGSADTQG 2035  
Db 1921 gissawanasmpqavipgrsmvtgciathnggavagvGLSKLSNDGQVFKINGSADTQG 1980  
Qy 2036 HVGAAGVAGGFHF 2047  
Db 1981 hvgaavaggfhf 1992  
RESULT 4  
AAB69137  
ID AAB69137 standard; Protein; 1992 AA.  
XX AAB69137;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE M. catarrhalis M56 200kDa protein in pKS348 SEQ ID NO:13.  
XX  
KW Moraxella catarrhalis strain Q8; major outer membrane protein;  
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection.  
XX  
OS Moraxella catarrhalis.  
XX  
PN WO200107619-A1.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-CA00870.  
XX  
PR 27-JUL-1999; 99US-0361619.  
XX  
PA (CONN-) CONNAUGHT LAB LTD.  
XX  
PI Loosmore SM, Sasaki K, Yang Y, Klein MH;  
XX  
DR WPI; 2001-159722/16.  
DR N-PSDB; AAF59106.



XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
PT useful in protective vaccines and for diagnosis  
PS Claim 1; Fig 8A-V; 247pp; English.  
XX  
CC The present invention describes an isolated and purified nucleic acid (I)  
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.  
CC The 200 kDa outer membrane protein (II) has antibacterial activity and  
CC can be used in vaccines. (II), and its truncated versions, are used as  
CC immunogenic compositions and vaccines to protect against M. catarrhalis  
CC infections, particularly otitis media in humans. (II) is also used as  
CC antigen in immunoassays for detecting specific antibodies (Ab), and to  
CC generate Ab. (I) are used for recombinant production of (II) and its  
CC fragments are used as probes for identifying/cloning 200 kDa protein  
CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
CC (I) makes possible production of large amount of recombinant immunogens.  
CC Expression of truncated versions of (II) reduces toxicity of the protein  
CC towards the Escherichia coli host. The present sequence represents the  
CC M. catarrhalis M56 200kDa protein in pKS348, which is given in the  
CC exemplification of the present invention.  
XX  
SQ Sequence 1992 AA;

Query Match 97.3%; Score 10021; DB 22; Length 1992;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1991; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 56 VIGATLSGSAAYAKQDKTKHTAIGEQNPQRSGTAKADGPAIAIGENANAQGGQATAIGS 115  
DB 1 migatlsagsayakdktkhtaigeqnqprsgtakadgdraiaigenanaqggqalaigs 60  
QY 116 SNTVNGSSLDKICTGATGQESTAIAGDVKASGDASTAIGSDDLHLLDQHCPKPKGTL 175  
DB 61 snktvngssldkictgatgesiaigdvkaskgdasiasgdsdlhldqhnkpkpgtll 120  
QY 176 INDLINGHAVLKEIRSKDNVRYRTTASGHASTAVGAMSAYAQGHFSNAPGTRATAKSA 235  
DB 121 indlinghavlkeirsskdnvryrttasghastavgamsayagghfsnafgtratak 180  
QY 236 YSLAVGLAATAEGOSTAIGSDATSSLSGAIAGAGTGAOLQGSIALGQGSVVYTSQDNNS 295  
DB 181 yslavglataegostiaigsdatslsigalagagtraqlqgsialgqgsvvvtqsdnns 240  
QY 296 RPAYTPNTQALDPKFOATNNTKAGPLSIGNSIKRKTIINYGAGVKNKTDAVNVAQLEAVVK 355  
DB 241 rpaytpntqaldpkfoatnntkagplsigsnsikrktiinygagvknktdavnvaql 300  
QY 356 WAKERRITFQDDNSTDVKTGLDNTLTIKGAETNALTNNIGVVKRADSGLKVLAKT 415  
DB 301 wakeritfsgddnstdvktglndtltikgaetnaltndnigvvrkadsnslkvlakt 360  
QY 416 LNNTEVNTTTLNATTNVKVGSSSTTAELLSLSLTTPQNTQSQTSTKTVYGVNGVKFT 475  
DB 361 lnnltevnttlnattnvkvgssttaelldseltfpqntqsqstsktvynvgvkft 420  
QY 476 NNAETTAAGTRITRDKIGFARGDVDEKQAPYLDKQKQLKVGSVVAITONGIDAGNKKI 535  
DB 421 nnaettaagtritrtdkigfargdvdeqapylkqkqlkvgsvaltdngidagnk 480  
QY 536 SNLAKGSSANDAVTLEOLKAAKPTLNAGATSVTPTEISVDKSGNVTAPTNYIGVKVTE 595  
DB 481 snlakgssandavtleolkakptlnagatstvptteisvdksnvtaptnyigkvte 540  
QY 596 LNSDGTSDKFSVKGSGNNLSLVAEHLASLYLNEVNRNTADSALQSFYTKEDDDANAITV 655  
DB 541 lnsdgtldkfsvkgsgnnslvtaehlaslylnevnrntadsalqsfytkedddanait 600  
QY 656 AKDTTKNAGAVSIILKKGKNGLTVAATKDGTVTFGLSQDSGLTITGKSTLNDGLTLVXDTN 715  
DB 601 akdttknagavsiilkgkngltvatkdgtvtfqlsqdsqslitgkstlndngltvkdtn 660

QY 716 EOIQVANGIKFTNVNNGPSTGIANTARITRDKTGFAGSDGAVDTNKPYPYLDQDKLQVGN 775  
DB 661 eqiqvngikftnvngnpgstgiantaritrdkifagsdgavdtnkpyldqdklqvgn 720  
QY 776 VKITNTGYNAGKATJGLSPTLPSIADOSSRNIELGNTIQDDKXNAASINDILNTGNL 835  
DB 721 vkitntgynagkatjglsptlpsiadossrnieigntiqddkxnaasindilntgfnl 780  
QY 836 KNNNPIDFVSVDIVDFANGNATTATVTHDPANKTSKVYDVNVDDTTHIHTGTDDNNK 895  
DB 781 knnpidfvsydivdfangnattatvthdpanktskvvydvnnvddtthihtgtdnnk 840  
QY 896 LGVKTTLNKTSAANGTATNFVNSDEDALVNAKDIAENLNTLAKEIHTTKGTADTALQ 955  
DB 841 lgvkttlknktsaangtntfnvnsdedalvnaadiaenlntlakeihthtkgtadalt 900  
QY 956 TPTVKKVDENNADDANATTVGOKNANNOVNTLTILKGENGLNKTDKNGTVTFGINTTSG 1015  
DB 901 tptvkkvdennaddanaltvgknannovntltikgenngnltikngnltkdkngtvtfginttag 960  
QY 1016 LKAGKSTLNDGGLSINKPTGSEQIQVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFT 1075  
DB 961 lkagkstlndgglsinkptgseqiqvgadgvkfakvnnngvvgagidgttritrdeigt 1020  
QY 1076 GTNGSLDKSKPHLSKDXINAGCKKITNQSGETAQNASHDAVTGGKIYDLKTELENKISST 1135  
DB 1021 gtngsldkskphlskdxinagckkitnqsgelagshdavggkiydlktelelenkisst 1080  
QY 1136 AKTAONSLHEFSVADBOGNFTVSNPYSYDTSKTSDDVITFACENGITTKVANKGVVRVGI 1195  
DB 1081 aktaonslhefsvadeognnftvsnpyssydtstddvitfagengitckvankgvrvgi 1140  
QY 1196 DQTKGLTTPKLTAVGNNGKGIVIDSQNGQNTITGLSNTLIANTVNTNKGSVRTTEQGNIIKD 1255  
DB 1141 dqtkgltpkltvgnngkgividsqngqntitglsntliantvntndkgsrvrtteqgnnikd 1200  
QY 1256 EOKTRAASIVDVLISAGFNQGNCEAVDFVSTVDTNFAOGNATTAKVTYDDTSKTSKVYV 1315  
DB 1201 eoktraasivdvlisagfnlqgnceavdfvstvtvtnfadgnattakvtyddtsktskvvy 1260  
QY 1316 DVNVDTTIEVKDKLIGVKVTTTTSTGTGANKFALSNOATGDALVKASDITVAHLNLTSGD 1375  
DB 1261 dvnvdttievkdklignvkvttstgtgankfalsnqatgdalvkasdlvahlnltlsgd 1320  
QY 1376 IQTAKASQANNSAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKKAKDKLVAQAQTP 1435  
DB 1321 iqtakasqannsagvydadgnkvlydystdnkyqakndgtvdktkkavakdklvaqaqcp 1380  
QY 1436 DGTLAOMNVKSVINKBQVNDANKKQGINEDNAPFKGLEKAASDNKTNAAVTVGDLNAYA 1495  
DB 1381 dgtlaomnvksvinkbqvndankkqginednapfkglekaasdnktnaavtvgdlnaya 1440  
QY 1496 QTPLTAFAGDTGTTAKLGETLTIKGGQTDNTNKLTDNNIGVWAGTDFGTVKLAKDLNLNS 1555  
DB 1441 qtpltfagdtgttaklgetlтикggqtdntnkltdnnigvvgagdtftvklakdlnlns 1500  
QY 1556 VNAGGTFIKDDKGVSPFVDSGQAKANTPVLISANGLDILGGKVISNVGKGTDKTDAANVQQLN 1615  
DB 1501 vnaggtkiddkgvspfvdsgqakantpvlisangldilggkvisnvvgkgtkdtdaanvqqln 1560  
QY 1616 EVRNLLGLGNAGNDADGNQVNTIADIKDPNCGSSSNRTVIKAGTVLGGKGNNDTEKLAT 1675  
DB 1561 evrnllglnagnadnagnvntiadikdpnsgsssnrtvikagtvlggkgnndteklat 1620  
QY 1676 GGTQGVVDKGNANGDLSNVWVTQDKGSKALLATYNAAGQTNLTNNPAAIDRINEQ 1735  
DB 1621 ggtqgvvdkgndgldslsnvvtqdkgskallatynaagqtnltnnpaaidrlineq 1680  
QY 1736 GIRFFHVNDGNQBPVQGRNGIDSSASGKHSVAIGFOAKADGAAVAIGRQTOAGNQSIA 1795  
DB 1681 girffhvndgnqbpvqgrngidssasgkhsvalgfoakadgaavaigrqtagnqgsia 1740  
QY 1796 IGSNAQATGDSQSIAGTGTGNVNVAGKHSIGAIDPSTVKADNSYSVGNNGQFTDQTQDFGV 1855

Db 1741 igbnaqatgddsgiaigtgnvvagkhsgaigdpstvkadnsysvgnnqftdatqtdvfgv 1800  
 Qy 1856 GNNITVTESNVALGNSAISAGTHAGTQAKKSDCTAGTTTAGATGKVGKFGAGOTAVGA 1915  
 Db 1801 gnnitvtesnvalgnsaisagthagtkqaksdgtagtttagatgtvkfagqavga 1860  
 Qy 1916 VSVGASGERRIONVAGEVSATSDVANGSOLYKATQOSIANANELDHRTHQENKANA 1975  
 Db 1861 vsvgasgerrionvagevsatstdvngsqlykatqqsianatneldhrrihqenkana 1920  
 Qy 1976 GISSAMAMSPQAYIPGRSMVTGSIATHNGOGAVAVCLSKLSLSDNGQWVFKINGSADTQ 2035  
 Db 1921 gissamamspqayipgrsmvtgslathngggavavglslsdngqvwfkingsadtq 1980  
 Qy 2036 HVGAAGVAGGPHF 2047  
 Db 1981 hvgaavgagfhf 1992  
 RESULT 5  
 AAB69135  
 ID AAB69135 standard; Protein; 2053 AA.  
 XX  
 AC AAB69135;  
 XX  
 DT 24-APR-2001 (first entry)  
 XX  
 DE M. catarrhalis strain Q8 200kDa protein SEQ ID NO:9.  
 XX  
 KW Moraxella catarrhalis strain Q8; major outer membrane protein;  
 KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
 KW otitis media; detection.  
 XX  
 OS Moraxella catarrhalis.  
 XX  
 PN WO200107619-A1.  
 XX  
 PD 01-FEB-2001.  
 XX  
 PF 26-JUL-2000; 2000WO-CA00870.  
 XX  
 PR 27-JUL-1999; 99US-0361619.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Loosmore SM, Sasaki K, Yang Y, Klein MH;  
 XX  
 DR WPT; 2001-159722/16.  
 XX  
 DR N-PSDB; AAF59104.  
 XX  
 PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
 PT useful in protective vaccines and for diagnosis -  
 XX  
 PS Claim 1; Fig 4A-V; 247pp; English.  
 XX  
 CC The present invention describes an isolated and purified nucleic acid (I)  
 CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.  
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and  
 CC can be used in vaccines. (II), and its truncated versions, are used as  
 CC immunogenic compositions and vaccines to protect against M. catarrhalis  
 CC infections, particularly otitis media in humans. (II) is also used as  
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to  
 CC generate Ab. (I) are used for recombinant production of (II) and its  
 CC fragments are used as probes for identifying/cloning 200 kDa protein  
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
 CC (i) makes possible production of large amount of recombinant immunogens.  
 CC Expression of truncated versions of (II) reduces toxicity of the protein  
 CC towards the Escherichia coli host. The present sequence represents the  
 CC M. catarrhalis strain Q8 200kDa protein, which is given in the  
 CC exemplification of the present invention.  
 XX  
 SO Sequence 2053 AA;

Query Match 74.5%; Score 7675; DB 22; Length 2053;  
 Best Local Similarity 74.8%; Pred. No. 0;  
 Matches 1604; Conservative 134; Mismatches 219; Indels 186; Gaps 28;  
 Qy 1 MNHYKVTFNKATGTFMAVEYAKSHSTGGSCATGQVGSVCTLSFARIAALAVLVIGAT 60  
 Db 1 mnhykvtnkatgtfmaaveyakshstggscatgqvgsvrtlsfariaalavlvigat 60  
 Qy 61 LSGSAYAOCKDTHHTAIGEONQPRS--GTAKADGDRAITAIGENANAOGGQAIATGSSNK 118  
 Db 61 lngsayagdttk-lelqgtknkintllkgdlatgeaslafigslskagsgaiaigsxvp 119  
 Qy 119 TVNSSLKIGTDATGOESIAIGDVKASGASIAIGSDDLHLDDQHGNPKHPKG--TLIN 177  
 Db 120 dpnngsgnvgshakgnesiaigdvlaegdaslaigsddlyl-----pknlldkn 170  
 Qy 178 D---LINGHAVLKEIRSKDNDVKYRRTTASGHASTAVGAMSYAQHFNSAFGTRATAKS 234  
 Db 171 efhklhghheilkkigtstgdkikyrrtraggghastavgamsyagghfsnafgtataea 230  
 Qy 235 AYSLAVGLAATAEGOSTIAIGSDATSSSLGATLALCAGTAAQLQGSIALQGSVVTQSDNN 294  
 Db 231 ayslavglaaqatkgsslavgsnakanaafaataiggnkvnlgrgvaigfgsqilrdnn 290  
 Qy 295 S-RPAYTPNTQALDPKFOATNNTKAGPL-SIG-----SNSIKRKTIINVGAGVKNKTDVNV 347  
 Db 291 tdasayvplgkltadqykatrggdstdfisgnsmnnssirrkliinvagarddavnv 350  
 Qy 348 AQLEAVVAKWAKERRITFO--GDDNSTDVKIGLDNLTITKGAETNALTNDNNNGVKEADN 405  
 Db 351 aqlkiveela-nrkitfkgdgnnsverglntltikgdaqtnalteanlgvv--tdg 407  
 Qy 406 SGLKVKLAKTLNNLTENVNTTILNATTVKVGSSSSTTAELLSDSLTFQPTNGTSSTSKT 465  
 Db 408 nglkvklakeitgltsv-----satnkitvsnntnnnaelsggittfs-pitgk-tdkt 460  
 Qy 466 VVGNGVKFTNN-----  
 Db 461 vysidglkftndsnsiatkgttrtkkkkigfagtdngvdeskpyldneklkvgnatlsng 520  
 Qy 478 -----AETTAIGTRITPRDKIGFARGDGDVDEKQAP 508  
 Db 521 sltvnnttgnkqivgangikfatvannvantsatvgtariceekigfagtdngvdeqap 580  
 Qy 509 YLDKKQLKVGSAITIDNGIDAGNKKISNLAKSSANDAVTIEQLKAAKPTLNAGANGISV 568  
 Db 581 yldkerlkvrveitdsginaghnkigtltnglantdavitkikldakptinagdgisi 640  
 Qy 569 TPTEISVDKSGNVTPATYNYIGVKTTTELNSDGT--DKFSVKGSGTNNSLVTAELHASYL 626  
 Db 641 nsngdglvdsngnitplynisvkttklnsgtsgnnkfsvsnahdnslvtakdladyl 700  
 Qy 627 NEVNRTADSAQSFTVKEEDDDANAITVAKDTTNKAGAVSILKLGKNGLTVATKK-DG 685  
 Db 701 nkvnetaadspsfkvqngdnnn-naivgkdt--ngktfntklkngenvnntnratg 757  
 Qy 686 TVTFGLSQDSGLTICKSLTNNDGLTVKDTNEIQVANGCIKFTNVNGSPGPGIANTARI 745  
 Db 758 tvtfglsgnslgttkpklcvsgd-----  
 Qy 746 TRDKTGFAGSGCAVDNTKPYLDQDKLQGVNVTITNGINAGGKATIGLSPTLPSTADQSS 805  
 Db 786 vieqv-psadg-----nstknllkigsptlpsiaspsg 817  
 Qy 806 RNIEGNTIQDKDKSNAASINDILNTGPNLKNNNRPIDFVSTYDIDVFANGNATTATVTH 865  
 Db 818 rnialgntieekdknsaasiddvlnaglnknkgdkdfvstydytdvfdignattatvty 877  
 Qy 866 DTANKTSKVVDVNVDDTTIHLTGTDNNKLGKVKTKLNKTSANGNTATNFVNSDEDA 925  
 Db 878 deandtskvaydvnvdektietlgtgngkklgkvktikitetstgn-attf--stddhda 934

QY	926	LVNADIAENLNTLAKEIHHTTGTADTALQTFTVKKVDENNADDANALITVCKNANQV	985
DB	935	lvkasdiagnlntlaeeihhtktgtentalqftfvkvdendkaddnaitvgkdgtskv	994
QY	986	NTLTLKAGNCLNLTDKNGTVFNGINTSGLKAGKS-TLNDGGLSIKKNPTSEQIQVGAD	1044
DB	995	ntlklkgngldlktokdgvfngintqgglkagdstltnnglsiktasneqilvgad	1054
QY	1045	GVKFAKVNNGVVGAGIDGTTTRTRDEIGFTGNSLDKSKPHLSKDGINAGGKKITNTQ	1104
DB	1055	gvkfamv-nngvvvgagldgttrtrtrdeigtgngslkskphlskdginaggkkitnlq	1113
QY	1105	SGEIAQNSHDAVTCGKIYDLKTELENKISTAKTAQNSLHEFSVADEQGNFTVSNPYSS	1164
DB	1114	sgeiaknshdavggklydlkteleknistaktagnslhefsvadeqgnftvsnpyss	1173
QY	1165	YDTSKTSDVITFAGENGITTKVKNKVVRVIGIDQTKLTPKLTGVNNGKGIIVDSQNGQ	1224
DB	1174	ydtsktsdvitfagengittkvknkvvrvgidqtkltpkltcvgnngkgiivnsqngq	1233
QY	1225	NTTGLSNTLANVTNKGSVRTTEOGNIKDEDKTRAASIVDVLISAGFNLOGNGEAVDFV	1284
DB	1234	nttfglsntlanvtnkgsrvrtteqgnilkdedktraasivdvlisagfnlogngeavdfr	1293
QY	1285	STYDTVNFADGNATTAKVTYDDTSKTSKVYDVNVDDTTIEVKDKKLGKVTTLTSTGTG	1344
DB	1294	stydtvnfngntttakvtyddtsktskvvydvnvddttievdkklgkvtltstgtlg	1353
QY	1345	ANKFALSQAATGDALVKASDIVAHLNLTSGDIOTAKGASQANNSAGYVDADGNKVYDST	1404
DB	1354	ankfalsnqatgdalvkasdivahlntlsdgiotakgasqannsagyvddadgnkvlydst	1413
QY	1405	DNKYQAANDGTVDKTEVAKDLVAAOQTPDGTLAQMNVKSVINKEOVNDANKKOGINE	1464
DB	1414	dnkyqaandgtvdktevakdklvaaoqtpdgtlaqmnmvksvinkeovndankkagine	1473
QY	1465	DNAFVGLKLEAASDNKTKNAAVTVGDLNVAQTPLTTFAGDTGTTAKKLGETLTIKGGQTD	1524
DB	1474	dnafvglekaasdnktnaavtvgdlnvnaqtplttfagdtgtttakkklgetltikgsgtd	1533
QY	1525	TNKLTDNIGVAGTGTFTYKLAIDLNLASVNAAGTPIIDDKGVSFVDSGSOAKANTPVL	1584
DB	1534	tnkltddnigvagtgtftvyklaidlnlasvnaagtpiiddkgsfvdsosgsoakantpvl	1593
QY	1585	SANGLDLGGKVISNVGKGTDKTDAAVQQLNEVRNLGLGNAGNDADGNQVNIADIKKD	1644
DB	1594	sangldlggkvisnvgkgtktdaavqqlnevrnlgl---gndnadgnqvnliadikkd	1650
QY	1645	PNSGSSNRVTIKAGTVLGGKGNNDTEKLTATGGTQGVGDKDGNANGDLSNVWVKTKDGS	1704
DB	1651	pnsghssnrvtikagtvlggkgnndteklataggvgvgvdkdgnangdlsnvwvktkdgs	1710
QY	1705	KKALLATYNAAGQTNLYTNPNPAEATRINEQIRFFHVNDCNQBPVQGRNGIDSSASGK	1764
DB	1711	kkallatynaagqtnvlytnpnpaetrineqirffhvnndcnqbvpvqgrngidssasgk	1770
QY	1765	HSVAIFQAKADGEAAVAIGRQTOAGNOSIAIGDNAOATGQOSTAIGTGNVVAGKHSQAI	1824
DB	1771	hsvalgfqakadgeaavaigrtqdegnqsiagldnaqatgqdsiaigtgnvvagkhsqai	1830
QY	1825	GPSTVTKADNSYSVGNNOFTDATQTOVFGVGNNTVTFTESNVALGSNSAITSAGTHAGTQ	1884
DB	1831	gpbstvkaadnsysvgnnoftdatqtdvfgvgnntlvtesnvalgsnsaitsagthagtq	1890
QY	1885	AKKSDCTAGTCTTTAGATGTGTVKFGAGTAVGAVSVGASGAERRIQNVAAAGEVSATSTDAVN	1944
DB	1891	akksdgtagtttagatgtvkvfgagtavgavsvgasgaerriqnvaagevsatstdavn	1950
QY	1945	GSQLYKATQSISANATNELDHRHQHENKANAGISSAMAMASNPQAYIPGRSMVTGGTATH	2004
DB	1951	gsqlykatqtsianatneldhrhqnknkanagissamamaspqayipgrsmvtggath	2010

2005 NGQGAVALGSLKSDNGQWVPFKINGSADTQGHVGAAYGAGFHF 2047

2011 ngqavavglsklsdngqwwfklngsadtqghvgaavagfhhf 2053

RESULT 6

AAB69136

ID AAB69136 standard; Protein; 2314 AA.

AC AAB69136;

XX 24-APR-2001 (first entry)

DE M. catarrhalis les1 200kDa protein SEQ ID NO:11.

KW Moraxella catarrhalis strain Q8; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

OS otitis media; detection.

XX Moraxella catarrhalis.

XX WO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loogmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX N-PSDB; AAF59105.

PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

PT useful in protective vaccines and for diagnosis

XX Claim 1: Fig 5A-Y; 247pp; English.

XX The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II), and its truncated versions, are used as immunogenic compositions and vaccines to protect against M. catarrhalis infections, particularly otitis media in humans. (II) is also used as antigen in immunoassays for detecting specific antibodies (Ab), and to generate Ab. (I) are used for recombinant production of (II) and its fragments are used as probes for identifying/cloning 200 kDa protein genes from other strains, and for diagnostic detection of M. catarrhalis. (I) makes possible production of large amount of recombinant immunogens. Expression of truncated versions of (II) reduces toxicity of the protein towards the Escherichia coli host. The present sequence represents the M. catarrhalis les1 200kDa protein, which is given in the exemplification of the present invention.

XX Sequence 2314 AA:

Query Match 64.0%; Score 6593.5; DB 22; Length 2314;

Best Local Similarity 60.7%; Pred. NO. 2.3e-282;

Matches 1466; Conservative 149; Mismatches 330; Indels 471; Gaps 48;

QY 1 MNHYKYVFNKATGTMAVAEYAKSHSHSTGGSCATGQVGSVCVTLSPARTAAALVLVIGAT 60

DB 1 mnhykyvfnkagctgmavaeacskshsgssstaggvgspsvtrtrvatlaillvgat 60

QY 61 LSGSAYAQKDKTHIAITGEONQPRRSCTAKADGDRATIGENANAGGQAIATGSSNKTV 120

DB 61 lngsayaq---nnsklafgttg---nndnasasneaisalgsalakahanqalaigsgkdpd 115

QY 121 -NGSSLDKIGTDTGQESIAIGGDVKAAGSDASIAIGSDDLHLILDQHGNPKHKGTLINDL 179



Thu Sep 13 14:18:50 2001

```
QY 1972 KANAGISSAMAMPMQAYIPGRSMVTGGIATHNGQGVAVGLSKLSDNGQVWFKINGSA 2031
DB 2239 kanagissamamampqayipgrsmvtggtiathngqgvavglsklsgdngqvwfkinga 2298
QY 2032 DTQGHVGAAGAGPFH 2047
DB 2299 dtqghvgaavgaagf 2314

RESULT 7
NA99393
ID AAR99393 standard; Protein; 2353 AA.
XX
AC AAR99393;
XX
DT 15-JAN-1997 (first entry)
XX
DE Haemophilus adhesion protein HA2.
XX
KW Haemophilus adhesion protein; HA2; hsf protein; vaccine.
XX
OS Haemophilus influenzae type b strain C54.
XX
PN WO9630519-AL.
XX
PD 03-OCT-1996.
XX
PF 22-MAR-1996; 96WO-US04031.
XX
PR 24-MAR-1995; 95US-0409955.
XX
PA (UYSL-) UNIV ST LOUIS.
XX
PA (UNIW) UNIV WASHINGTON.
XX
XX Barenkamp SJ, St Geme JW;
XX
XX WPI; 1996-455364/45.
XX
XX N-PSDB; AAT41476.
XX
XX Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in
XX vaccines against H. influenzae infection.
XX
XX Claim 5; Page 66-73; 120pp; English.
XX
XX Haemophilus adhesion protein HA2 (AAR99393) is associated with the
XX formation of surface fibrils involved in adhesion to various host
XX cells; it is also referred to hsf (Haemophilus surface fibrils).
XX Its amino acid sequence was deduced from a genomic DNA clone
XX (AAT41476) derived from Haemophilus influenzae type b strain C65.
XX Large quantities of recombinant HA2 can be produced in transformed
XX prokaryotic or eukaryotic host cells, for use in vaccines against
XX H. influenzae infection.
XX
XX SQ Sequence 2353 AA;

Query Match 10.9%; Score 1123; DB 17; Length 2353;
Best Local Similarity 23.0%; Pred. No. 1.6e-41;
Matches 607; Conservative 318; Mismatches 838; Indels 874; Gaps 119;

QY 1 MNHIIYFNKATGTFMAVAYAKSHSTGGSCATGQVGSVCTLSFARIALAV----- 54
DB 1 mnkifnviwmvtqtwwvseilrthtk--rasatvetavlatlilafvcqanadedeel 58
QY 55 -LVIGNTLGSAYAKDKTQKHIAIGEONQPRRS-----GTAKADGDRAIAIGENAN 104
DB 59 dpvrtapvlsfhsdegt-----gekevtenwnwgiyfdnkvla-gaitlkagdnik 112
QY 105 AQCGQAIATGSSNKVTNVS-----SLDKTGTDTAT--GQESIAIGGDVAKSGD-ASIAIGSD 157
DB 113 ik-----qntdesnassfysikkldtlttsvateklsgf-----angdkvdtlsdan 161
```

```
QY 158 DLHLLDHGHNPKHPKG--TLINDLINGHAVLKEIRSKONDVKYRRT----- 202
DB 162 glklaktgngvnhngldstlpdavnrtgvlss--sstfpndvkekraatvkdvlmagwnl 220
QY 203 ----TASGHASTAGVAMSYAQGHFSNAFGTRATAKSAISLAAGLAATAEGOSTIAIGSDA 258
DB 221 kgaktagnvesvdlvsaynnvefit--gdknt-----ldvltakengkttevkftpk 272
QY 259 TSSSLGAIALGAGTRAQLOGSIALGOGSVVTSQDNNRPPAYTPNTQALDPKFOATNNTKA 318
DB 273 ts-----vikekdglftgkenndtnkvtant-----atdntde 306
QY 319 GPLSIGNSIKRKIIINVAGVKNKTDVAVVQAQLEAVVWAKERRITTFQGDNDST----- 371
DB 307 gnglvtakavidavnkagrvktttangngdfatv--asgtnvfesgogdtasvckdt 364
QY 372 -----DVKIGLDNT--LTIKGG----- 386
DB 365 nnglvtkydakvgdglkfdskkivadtaltvtgkvaekadeddkklvnagdlvta 424
QY 387 -----AETNALTDNNI--GVVKEAD-----NSGLKVK-----LAKTLNN 418
DB 425 lgnlswkakaeadtdgallegiskdgvkagetvtfkagknlkvkqdgangftysiqdaltg 484
QY 419 LTEVNTTLNATTTVKVGSSTTAELLSDSLTFQPTNGSQSSTKTV--YGVNGVKKFTNN 477
DB 485 ltsi---tlggttn---ggndaktv--lnkdglitpagnnggtctntisvtkdkgkagnk 537
QY 478 AETTAAG-----TTRITRDKRIGFARDG--DVDEKQAPYLDKOLKVGVS 520
DB 538 altnvasglrayddanfvlwnnsatdlr--hvedaykglmlnekna---nkqplvtdat 593
QY 521 AITIDN-----GIDAGNKKISNLAGSSANDAVTTEQLKAAKPTLNAGAGISVTPTE 572
DB 594 aatvgdlrklgwwvstkngtkeesn-----gvkqadeviftgagaa----- 634
QY 573 ISVDAKSGNVTAPTNYIGVKTE---LNSDGTSDKFSVKGSGTNNLSVTAHLASYLNE 628
DB 635 -tvtsksen--gkhtltvsvaetkadcglekdgtiklkvdnqntdnvltvgn----- 684
QY 629 VNRTADSLQSFVKE--EDDDANAITVAKDTTKNAGAVSILKLGKNGLIVATVKKDGTV 687
DB 685 -ngtavtkggefvtktgatdargkvtv--kdatandakkkvatvk-----dvatalnaaa 737
QY 688 TP-----GLSQDSGLTIGKSTLNNDLTVKDR-----NEQIQVGANGIGFT----- 728
DB 738 tfvktenlttsidednptdngk-----ddalkagdtltfkagknlkvkdgnitfdlkn 793
QY 729 -----NVNGSNPFGTGIANTARI---TRDKIGFAGSDGAVDTNKPYPLODQKLQ 772
DB 794 levktakvadtltigntptgttatkpnitstadglntfa----- 834
QY 773 VGNVKTNTGINAGGKAI--TGLSPTL--PSIADQSSRNIELGNTIQDKDKSNAASINDI 828
DB 835 -----ketadasgsknylkgiatlttpepsagaks--hvdl--nvdatkksnaaaiedv 885
QY 829 LNTGENLKNNNPIDFVSTVIDVFANGNATATVTHDTANKTSKVYVDNVDDTTHIT 888
DB 886 lragvniqgnnnydyatydytvtvncddstgttvt-----vtqkad 927
QY 889 GTDDNKKLGKVTTKLKNKTSANGNTAT-----NFVNNSDEED-----ALVNAKDAIEN 935
DB 928 gkgadvkigaktsvik--dhngklftgkldkdnngatvseddgdgkdtgtgvlcaktvda 985
QY 936 LNTLAKEIHTTKGVADTALQTFVKKVDENNADANAITVGO-----KNANNOVNTLTLLK 991
DB 986 vnksgwrvtgegaetaetgatav-----nagnaetvtsglsvnfkngn--attatvs 1034
QY 992 GENG--LNKIKTDKNGTVFNGINTVSGLKAG-----KSTLNDGGLSIKNPTSEQIQ-- 1040
DB 1035 kdnglnvnykd-----vnnvgdglkigdkkivadtltltvgkvsypaganannnn 1086
QY 1041 ---VGADGVKFAKYN-- 1053
```

Db 1087 kklvnaeglatlannlswtakadydgesegetdevkagdkvtfkagknlkvkqsek 1145  
QY 1054 -----NGVVGAGIDGT-----TRITREICGTGTNGS-----LQSKPH-----LS 1089  
Db 1146 dftyslqdtlgtlscitlgtangrndgtvinkdgtltilangaaagtddasngntlsvt 1205  
QY 1090 KGINAGCKKITNIQSG-----ETAQNSHDVATGCKIYDLKYTELENKI-----STAK 1137  
Db 1206 kdgisagkeltvksalktykdtqntade-tgdkfhaavknaevfvgkngatvsak 1264  
QY 1138 TAONSLH-----EFSVADQGNFTVSNPSSYSDTSK----- 1169  
Db 1265 tdngkhtvidvaeakvgdgldkdkiklvndtdgmllctvdtkgaaavakgefna 1324  
QY 1170 -TSDVITFAGENG-----ITTKVKNKVVRVIGIDQTKLT----- 1202  
Db 1325 vttdattagtnanergkvvvksgngatatetdkkvtatgdvavakaindaatfvkvvendd 1384  
QY 1203 -----TPKLT 1207  
Db 1385 satidsptddgandalagdkdtlalkagknlkvkrdknltfalandsvksatvsakls 1444  
QY 1208 VGNNNGK-----GIVIDSONGNT-----INGLSNTLANVTNDKGSVRTTEQGNLIK 1254  
Db 1445 lgtngkvnitsdtkglnfakdkstgtdgdaninhnglgiastltdtllnsgat-tnlgngit 1503  
QY 1255 DEKTRAASIVDLASGNLQSG-----NGEAVDFVSTYDVTNFDAGNATTAKVTYD- 1305  
Db 1504 dnekkraavkvlnagvnrvgkpasannqvenidfvaitydvtvfygdkdtsvtves 1563  
QY 1306 -DFTSKTVKVVYDNDVDTTIEVKDKKLGKVTYTTSTSG-----TCAN-KFALSNOAT----- 1355  
Db 1564 kdngkrtev-----kigatsvikhngklftgkelkdannngvtvtet 1607  
QY 1356 -----GDALVKASDIVAHLNLTSGDIOTAKGASQANNSAGYVD-----ADGNKVIYDS 1403  
Db 1608 dgkdegngltakavidavnkagrvvkt-gangqnddfatvasgtnvtfagngngtaev 1666  
QY 1404 TDNKYQAKNDG---TVDKTKEVAK-----DKLVAQA-----QTPDGTIAQMNKSVINKEQ 1452  
Db 1667 t-----kandgsitvkynkvagdklkdgkivadtvtltvaagkvktapn----- 1711  
QY 1453 VNDANKKQGINEDNAFVKGLEKAASDNKTKNAAVTGDNLNAVAQTPLTFAGDGTGTTAKKL 1512  
Db 1712 -ngdgkk-----fvdasgladalnklshta-tagkegtgevdpsansagg-----evka 1757  
QY 1513 GETLTIKGQDTNKLTDNNIGVVAGTGDGFTVKLAKDLTNLSV-----NAG-----GFKI 1563  
Db 1758 gdkvtftag-----dnlkikqsgkdfyslkkelkdltsvefkddanggtgsestki 1808  
QY 1564 DDGVSFVDSGQA-----KANTPVLASGLDLGKRVISNVGKGT-----DTDA 1609  
Db 1809 tkdglitpangagaagantantitsvckkgisagkavtnvsvglkkfgdghtlangtva 1868  
QY 1610 NVQQL--NEVRNLLGLGNAGNDN-----ADGNQVNIAD-----IKKDPNSGS----- 1649  
Db 1869 dfekhydnaykdkltlnldekadnnpvtadntaatvdlrglwgvisadkttgepnqeyna 1928  
QY 1650 ---SSNRTVTKAGT-----VLGG-----KGNNDTEKLIATGG-----IQVG 1681  
Db 1929 qvrnanevkfsgnginvsgktlgtngtrvtfelakgevwksneftvknadgsetnlkvvg 1988  
QY 1682 -----VDKDNANGDLSNVWVKQT-----DGSKKALLATYNAAGQNTYLTNN 1724  
Db 1989 dmyyskedipatskpmgtg---ktekykvengkvvsangsktevtltnkqsg---yvtgn 2042  
QY 1725 PAEALDRINQGIREFHVNDGNQBPVVQGRNGIDSSASGKHSVAIGFOAK----- 1774  
Db 2043 --qvadalaksgfel-----gladaaeakfaesakdkqlskdkaetv 2084  
QY 1775 -ADGEAAVATGROPTAQGNQSIATGDNAQATGQ-----SIAIGTGVNVVAKHSGAIG 1825

Db 2085 nahdkvrfanglnktkv-----saatvestdangdkvtttfvtvkdvelpltcqiy---ntdang 2138  
QY 1826 DPSTVKADNSYSVGNNNQFTDATQTDVEFGVNITVTESVSVALGSNSAISAGTHAGTQA 1885  
Db 2139 nkivkkadgkw-----yelnadgtasnkvevtlgnvdangkk-----vv 2176  
QY 1886 KKSOGTAGTCTTTAGATGTVKGFAGOTAVGAVSVG-----ASGAERTONVAAG 1933  
Db 2177 kvtenagdkwyytnadgaadkktkgevsndkvstdekhhvrvldpnnqsgkgvvidnvang 2236  
QY 1934 EVSATSTDAVNGSOLY---KATOSTANATNELDRIHONENKANAGISSAMAMASMPQAY 1990  
Db 2237 eistatdaiingsqlayavakvtnliagvnnlekvkvkradagtsalaasqipgat 2296  
QY 1991 IPGRSMVTGGIATINGGAVAVGLSKLSDNGQWVFKINGSADTQGHVAAVAGAGPHF 2047  
Db 2297 mpgksmvaiaagssyggnglaigvrsisngkviirlsgttngsgktgvaagvgvqw 2353  
RESULT 8  
AAB23860  
ID AAB23860 standard; Protein; 2411 AA.  
XX AC AAB23860;  
XX DT 17-JAN-2001 (first entry)  
XX Haemophilus influenzae adhesin (Hia) protein from type c strain API.  
DE Hia; adhesin; Haemophilus influenzae adhesin; NTHI; infection; vaccine;  
KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;  
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;  
KW diagnosis; immunogenic; antigen.  
XX Haemophilus influenzae.  
OS W0200055191-A2.  
IN 21-SEP-2000.  
XX 16-MAR-2000; 2000WO-CA00289.  
XX 16-MAR-1999; 99US-0268347.  
XX (CONN-) CONNAUGHT LAB LTD.  
XX Loosmore SM, Yang Y, Klein MH;  
XX WPI; 2000-618897/59.  
XX N-PSDB; AAA92499.  
PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for  
PT use as antigens and vaccines and for treating Hemophilus influenzae  
PT infection  
XX Claim 1; Fig 24; 275pp; English.  
XX The present sequence represents a Haemophilus influenzae adhesin  
CC (Hia) protein from the type c Haemophilus influenzae strain API.  
CC Hia genes and proteins have antiinflammatory, auditory and antibacterial  
CC activities, and can be used in the production of a vaccine. An  
CC immunogenic composition comprising an Hia gene, a polypeptide encoded  
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
CC protection against disease caused by Haemophilus strains in a  
CC susceptible host, preferably a human. An Hia protein is useful as an  
CC antigen, in immunogenic preparations including vaccines, as a carrier  
CC for other immunogens, and in the generation of diagnostic reagents. Hia  
CC is useful for treating diseases caused by the infection of Haemophilus  
CC influenzae such as meningitis, epiglottitis, septicaemia and otitis  
CC media. Recombinant production of Hia favours high recovery of the  
CC protein compared to the low recovery of native protein from Haemophilus  
CC influenzae species. A truncated protein has a significantly higher  
CC amount of recovery than a full-length protein.

XX	Sequence	2411 AA;
SQ	Query Match	10.8%; Score 1109; DB 21; Length 2411;
	Best Local Similarity	23.2%; Pred. No. 6.8e-41;
	Matches	623; Conservative 306; Mismatches 845; Indels 910; Gaps 124;
QY	1- MNHIKVIENKATGTFMAVAEYAKASHSGSCATCGVSGVCTLSFARIAALAV-----	54
Db	1 mkiifvnmvtqtwwvseltrthtk--rasatvetavlatlifatvqanadedeel	58
QY	55 -LVIGATLSGSAYAKOKTKHIAIEQONPRRS-----GTAKADGDRARIAIGENAN	104
Db	59 dpvrvrtapvlshfsdkegt-----gekevtensnwgifyhknkvika-gaitikagdnlk	112
QY	105 AOGGOAIAIGSSNKTVNGSLDKIGIGDAT--GOESTAIIGDDVKASGD-ASIAIGSDDLHL	161
Db	113 ik-----qstnasftyslkldltsvateklisf---angdkvdltsdanglkl	161
QY	162 LDQHGNPKHPKG--TLINDLINGHAVLKEIRSSKONDVKYRPT-----	202
Db	162 aktgngvnhngldstlpdavnrtgvlss--sftpnndvakraatvkvlnagwnikgak	220
QY	203 TASGHASTAVGAMSYAQGHESNAFGNTRATAKAYSILAVGLAATAEGOSTIAIGSDATSS	262
Db	221 tagngvesdvlsaynnvefit--gdknt-----ldvltakengkttevkftpkts--	270
QY	263 LGATIALCAGTRAQLOGSIALGOGSVVTVQSDNNSRPAYTPNTQALDPKFOATWNNTRAGPLS	322
Db	271 -----vikekdglftgkenndntnkvtsnt-----atdntdegngl	306
QY	323 IGNSIKRKIIINVGAVNKTDAVNVQAQLEAVVKNKERRITTFQGDONSPT-----	371
Db	307 vtakavldavnkagrvkrttngngngdfatv--asgntvrfesgggttasvktcdngng	364
QY	372 -----DVRIGLDNT--LTIKGAETNALTDNNGVYKEAD-----NSGL	408
Db	365 itvkydakvgdglkfdsdkkivadtaltvtgkvae-----fakeddkklvnag-	415
QY	409 KVKIAKTLNNT-----EVNTTPT-----LNATYTKVKGSSSTTAELLSDSITETQPN	457
Db	416 --divtalgnlswkakaeadtdtdgalleglskdqevkag-----etvtf--ka	459
QY	458 GSQSTSTVYGVNGVKTNNAEATTAAGTTRITRDKTIGFARDGVDVDEKQAPYLDKKQKLV	517
Db	460 gnlkvkq-----dganfyslq-daltgltsit--lggtngngndaktv--inkdgl	509
QY	518 -----GSVAITIDNGIDAGNKISMLAKGSSANDAVTIEQLKAAKPTLN-----	561
Db	510 tpagnggtgtntisvtkd-gikagnkaintvasglrayddanfvlmsatdlnrhved	568
QY	562 AGAGTSVPTTEISVDKSGN-----VTAPT-----YNIGVKPT--TELSNDSGTSDKF	605
Db	569 aykgil-----lnlnknankqplvtstaataatvgdlrklgwvstknktkeesngkvkqad	622
QY	606 SVKSGCTNNSLVTA-----EHLASYLNEVNRNTADSAQSFVTKEDDDANAITVAKDPTT	660
Db	623 evlftgagaatvtsksengkhtit--vsvaetkadsgl-----ekdgdtklklvngnt	674
QY	661 KNAGAVSILKLGKNGNLYAVTKDGTWTFGLSQ--DSGLTIGKSTLNNDG-----LTVKDTN	715
Db	675 dnvlvt-----gnggtavtkggtfvtktgatdadrkvtvkdatadadkkvatkvkva	728
QY	716 EOIQOVGANGIK-----FTNVGNSNPGTGIANTRITRDKIGF--AGSDGAVDFTNKPYLDOO-	769
Db	729 tainsaatvfktenltsidnpldngkdallkagdtltfkagknkvlkrdngknitfdl	788
QY	770 --KLOVGNVKTNTWGINAGKAITGLSPTLPSI-----ADQS--SRNIELG	811
Db	789 aknlvktakvsdt-itigntptggttatpkvknitstadglnfakctadsgsknvylyk	847
QY	812 N-----TIQDKDKSNAASINDILNTGFNLKNNNPIDFVSTYDIDV	852

Db	848 giattltpepsagaksshvdlnvdatkknaasiedvlragwnlqngnnvdyvtydtvn	907
QY	853 FA-----	854
Db	908 ftdstgtttvtvtqadgkadvkigaktsvikdhngklftgkdkldannngatvseddq	967
QY	855 -----	859
Db	968 kdtgtglvtaktvldavvnksgwrvtggegataetgataavnagnaeetvtsvtnfkngnat	1027
QY	860 TATVTHDTANKTSKVYVDVNV-----DFTTTHLTG-----TDDNK	894
Db	1028 tatvskdnng--lnvkydvnnvgdglkgdkkivadtltltvtggykxvvpagansvnnnk	1085
QY	895 KL-----GVKTTKLKNTS-----ANGNT-----ATNFNVNNSDEED	924
Db	1086 klvnaegil-atalnnlswtakadyadesegeldqevkagdkvtfkagknkvlkvksekd	1144
QY	925 ALVNAKDI-----	940
Db	1145 ftyslqdtltgltsitlgtgtangrntgtvinkdglititlangaaegtddasngntisvtk	1204
QY	941 -----KEIHTTKGTADTALQTFTVKKVD-----ENNAD	969
Db	1205 dgisagnkeltnvksalktykdtqntagatqpaantaevakqdlvdltkpatgaangad	1264
QY	970 ----DANAITVQO-----KNANNOVNPTLTUKGNGNLNI--	998
Db	1265 akapdttaatvgdlrglgwvisakktadtqdkfhaavknane-----vefvgkngatvs	1320
QY	999 -KTDKNGVTGINTTSGLKAG-----KSTLNDGGLSIK--NPTSEQIOVGA-DGVKFAK	1050
Db	1321 akttdngkhtvtidvaea-kvgdglekdt--dgkiklkdvdcdgnnlltvtatkgasvak	1377
QY	1051 VNNNGVW--GAGIDGTRITRDEIGTGTNGS---LDKSRPHLSKDGINAGGKKITNIQ	1104
Db	1378 gefnavttdattagatqnanergkvvvkgsngatatetdk-----kkvatv-	1422
QY	1105 SGEIAQNSHDVATGKGIYDLTKTELENKISSIAKTAQNSLHESVADEQGNFTVSNPYSS	1164
Db	1423 -gdvaka lndaatfvkv-----enddsatl-----ddsptcdgan-----	1456
QY	1165 YDTSKTSVDITE--AGENGITTKVNGV---VRVGIDQTKGLTTPKLTGNNNGK-----	1214
Db	1457 -dalkagdtltikagknkvlkrdgnkntifalandsvksatvsdklsigtngknvnltsd	1515
QY	1215 ----GIVIDSQNGNT---ITGLSNFLANVTNDKSGVTRTEOGNIKDEDKTRAASIVDV	1267
Db	1516 tkglntfakgsktgddanihnglasltitdlinsgat-tnlgngitdnekkraasvkv	1574
QY	1268 LSAQFNLOG-----NGEAVDFVSTYDVFNFADGNATTAKTATVD--DTSKTSKVYVDV	1317
Db	1575 lnagvnrvgvtpasaannqvenldfvatvtdvtfvsgdkdttsvtveskdngtrtev----	1630
QY	1318 NVDDTTIEVKDKKLGVKTTTLTSTG---TGAN--KPAALSQAT-----GDALVKAS	1363
Db	1631 -----kigaktsvikdhngklftgkelkdannngvttvtetdgdgdegnglvtak	1678
QY	1364 DIVAHLNLTSGDTQAKGASQANNAGYVD-----ADGNKVIYDSTDNKYQAQNDG-	1415
Db	1679 avidavnkagrvktt-gangngndfatvasgtnvtfadngngtaevt-----kandgs	1731
QY	1416 -TVDKTKVEAK-----DKLVAQA---QTPDGTLAQNMNVKSVINKFQVNDANKQOINED	1465
Db	1732 itvkvynkvvadgliklkgdkkivadtltvtvadgkvtpgn-----ngdggk-----	1775
QY	1466 NAFVKGLEKAAASNKYNKAAVTVGDLNVAQOTPLTFAAGDTGTAKKLGETLTKGGQTDI	1525
Db	1776 --fvdasgladlnklshta-tagkegtgevdpanaagq-----evkagdkvtfkag----	1824
QY	1526 NKLTDDNNGVAGTGTGFTVKLAKDLTNLSNV-----NAG-----GTIKDDKGVSVFSSQG	1576



Db	1825	-----dnllkiksgkdfityslkkelkdltsvefkdanggtgsestkttkkdglittpanga	1870
Qy	1577	A-----KANTPVLISANGLDLGKVIISNNGKGTK-----DPTDAANVOOL--NEVRNL	1620
Db	1880	gaagantantvisvtkdgisagnkavtnvsgllkfgdghtlangtvadfekhydnaykdl	1939
Qy	1621	LGLGNAGNDN---ADGNOVNTAD-----IKKDPNSGS-----SSNRTVIKAG	1659
Db	1940	tnlidxgdndnptvadtaatvgdlrglwgwisadkttgpegnaynaqvrnanefksg	1999
Qy	1660	T-----VLGG-----KGNNDTEKLATGG-----LOVG-----VDKDGNA	1688
Db	2000	nginvgsktlngrtrvitfelakgevksnefvknadgsetnlvkvqdmymyskedldpat	2059
Qy	1689	NCDSLNVWVKTK-----DGSKKALLATYNAAGQTYNLTNPAIDAIRINEQGI	1737
Db	2060	skpmtg---ktekykvangkvsangsktevtltnksg--yyvtgn--qvadaiaaksgf	2111
Qy	1738	RFPHVNDGNQEPVVGVRNGRIDSSAGKSUSVAIGFOAK-----ADGEAAVAIGRQ	1786
Db	2112	el-----gladaaeaeakafesaakdkqlskdaetvnaahdkvrfangln	2155
Qy	1787	TOAGNOSIAIGDNAQATGDO-----SIAIGTCNVVAGKHSGAIGDPSTVVKADNSYSV	1838
Db	2156	tkv---saatvestdangdkvtftfvdktelptlqly---ntdangnkivkkadgkw--	2207
Qy	1839	GNNNQFTDATQDFVCGVGNITVTESNVALGCSAISAGTHACTQAKKSDGTAGTGTTTA	1898
Db	2208	-----yeinadgtasnkvtlgnvdangkk-----vvkvctengadkwytt	2247
Qy	1899	GATGTVKGFAGQTAVGASVSG-----ASGAERRIQNVAAGEVSATSDAVNGS	1946
Db	2248	nadgaadaktkgevsndkvtdekvhvridpnqngkvvldnvangeisatstdaings	2307
Qy	1947	QLY---KATQOSTANATNELDRIHQENENKANAGISSAMAMASMPQAVIPGRSMVTCGIAT	2003
Db	2308	qlyyavakgvtnlagqvnnlegkvkvkgradagtasalaasqlpqaumpdksmvaiaags	2367
Qy	2004	HNGQAGAVAVGLSKLSDNGQWFKINGSADTQCHVCAAVGACGFHF	2047
Db	2368	yqgqnglaigvrisdngkvkirlsgttnsggktgvaagvyqvw	2411
RESULT	9		
AAAB23856			
ID	AAAB23856	standard; Protein; 1104 AA.	
XX			
AC	AAAB23856;		
XX			
DT	17-JAN-2001	(first entry)	
XX			
DE	Haemophilus influenzae adhesin (Hia) protein from NTHi strain 29.		
XX			
KW	Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine		
KW	non-typeable Haemophilus influenzae; antiinflammatory; auditory;		
KW	antibacterial; meningitis; epiglottitis; septicemia; otitis media;		
KW	diagnosis; immunogenic; antigen.		
XX			
OS	Haemophilus influenzae.		
XX			
PN	WO200055191-A2.		
XX			
PD	21-SEP-2000.		
XX			
PF	16-MAR-2000; 2000WO-CA00289.		
XX			
PR	16-MAR-1999; 99US-0268347.		
XX			
PA	(CONN-) CONNAUGHT LAB LTD.		
PI	Loosmore SM, Yang Y, Klein MH;		
XX	WPI; 2000-618897/59.		
DR			

```

DR N-PSDB; AAA92495.
XX
XX Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating Hemophilus influenzae
PT infection
XX
XX Clalm 1; Fig 20; 275pp; English.
XX
XX The present sequence represents a Hemophilus influenzae adhesin (Hia)
CC protein from the non-typeable Hemophilus influenzae (NPHi) strain 29.
CC Hia genes and proteins have antiinflammatory, auditory and antibacterial
CC activities, and can be used in the production of a vaccine. An
CC immunogenic composition comprising an Hia gene, a polypeptide encoded
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
CC protection against disease caused by Hemophilus strains in a
CC susceptible host, preferably a human. An Hia protein is useful as an
CC antigen, in immunogenic preparations including vaccines, as a carrier
CC for other immunogens, and in the generation of diagnostic reagents. Hia
CC is useful for treating diseases caused by the infection of Hemophilus
CC influenzae such as meningitis, epiglottitis, septicaemia and otitis
CC media. Recombinant production of Hia favours high recovery of the
CC protein compared to the low recovery of native protein from Hemophilus
CC influenzae species. A truncated protein has a significantly higher
CC amount of recovery than a full-length protein.
XX
XX Sequence 1104 AA;
SQ
Query Match 9.98; Score 1016; DB 21; Length 1104;
Best local Similarity 27.9%; Pred. No. 3.4e-37;
Matches 373; Conservative 166; Mismatches 405; Indels 394; Gaps 55;
QY 869 NKTSKVYDVNVDD-----TTIHLTCTDONKKLGVKTKLNKTS-ANGNTA----- 913
DB 2 knifnvlvnnvtqtvvvseltrhtkcoasatvavatlatsacaennntsvtnglna 61
QY 914 ---NFNF-VNSSEDEALVNAKDIAENLNTLAKEHTHTKGTADTALQTFTVKVDENNAD 969
DB 62 ygdtnftnnsiadlekvhqdaykgllnlek-----dtnkssflva----- 104
QY 970 DANATVIG-----QKNAN-----NQVNTLTLLKGENGLNI-KTDKNGTVPFGI 1010
DB 105 dntaatvgnrlkglwvsskngtnkcsyqvkqadevlftgsgaatvsssskdgkhtltl 164
QY 1011 NTTSSLKAG-KSTLNDGSLTKNPTGSRQIQVGADGVKFAKVNNGVVGACIGDTTRTR 1069
DB 165 svtkgsfaevktdattgg-----qvnadrkvkkaeden---gaadv----- 202
QY 1070 DEIGFTGTNGSLDKSPHLSDKGINAGGKKTNTTQSGEIAQNSHDVATGGRIYDLKTELE 1129
DB 203 -----kkvatvk--dvakaindaatfvkvestdddie 232
QY 1130 NKISSTAKTAQNSLHEFSVADEQGNFTV---SNPYSYDPTSK7SDVITTFAGENGITKV 1186
DB 233 ngasgknettdqal-----kagdtlltkagknlkakldqngksvtfalakkldvtsak 285
QY 1187 NKGVVVRGIDQTK-----GLTPPKLTVCNNNGKGVIDSQNGONTITGLSNTLANVT 1238
DB 286 vsdklsigkdcnkviditsdangllak--tgngng-----qngnvhlinglaatl---- 332
QY 1239 NDKGSVRTTEOGNIIKDEDKTFAASIVDVLGAGFNLOGEAVDFVSTYDVTVNFADGNAT 1298
DB 333 tdtlgtmtqasngvavqgnhraasvadvlnagwniigngasvdfvnydaytdvfgngtnt 392
QY 1299 TAKVTYDDTSTKTSKVYVD-----VNVDDTTFIEVK-----DKKLG--- 1332
DB 393 nvnvttdtahkktlrvrdvtglpvyvtedgktvvkvdnkcyekakqdgsgdmkdvke 452
QY 1333 -VKITTTLTSTGTGANKFALSNOQTGDALVKA5DIVA--HLNTLSGDIQTAKAGSOANNSA 1389
DB 453 laktkvklvsaggnpvklsnvaeg---teendavsfkqlkalg-ekqvtlitasnayang 508
QY 1390 GYVDADGKNKV1-----YDSTDNKYYQAK--ND-----GTVDKTKFEVAKNKIV 1429

```

```
Db 509 gn-daggkatqtlnglnfkfstgdgellnikvendtvtlfpkkgsv-qvgedgkatiq 566
QY 1430 AQAOPTDGTIAQNNVSVINKBOVDANKKOGINEDNAFVKGLEKAASDNKTKNAAVTVG 1489
Db 567 ngkttdglvaseel-----veslnklgwkvgdkdgs-----g 600
QY 1490 DLNAVOTPLTFAGDTGTTAKKLGELTTLTKGGOTDNTKLTDDNIGVVAGTDGFTVKLAKD 1549
Db 601 eldgasnetlvksgdkvtl--kagenlkvk--qdgcn-----ftyalkde 641
QY 1550 LTNLSV-----NAGTKIDDRGVSVFSSG-----QAKANTFVLSANGLDGGKV 1595
Db 642 itgvksvefdtangangastkiktdgltitsangangaaatadadkikvasdgisagnka 701
QY 1596 ISNVGKGTDTDAANVQOLNEVENLL-----GLGNAGNDNADGNQVNIADIKDPN 1646
Db 702 vknvsglkkfgdanfnlptssadnltkgydaykgltnldexgdkqtlitvad-----n 756
QY 1647 SGSSNRTVIKAGTVLGGKGNNDTEKLTATGGIQGVGDKDGNANGDLSNV-WVKTKDGSK 1705
Db 757 t-----aatv-----gdrlglgwv-isadktt 777
QY 1706 KALLATYNAAGQNTLYTNPAEIDRINEQGIREFHVNQNGQEPVVGQNGIDSS---AS 1762
Db 778 gelnkeyna-----qvrnanevfkfsg-nglihvsgktvn 810
QY 1763 GKHSVAITGFOAKGDEAAVAIGROTQAGNOSIATGDNAAQTGDSQTAIGTGNVAVACKHSG 1822
Db 811 grrait--felakdena-----lafgyskalrdntvaigcgvvnaeksg 854
QY 1823 AIGDPSTV--KADNSVSGNNOFTDATQTDVFGVGN----- 1858
Db 855 afgdpnyiedkaggsyafgdnrit-skntfvlngvngakngdvdttvtvkdkdkg 913
QY 1859 ---ITV-----TESNSVALGSNSAISAGTHAGTQAKKSDGTAGTTTAGATGTVKGFAG 1909
Db 914 ettvtvpkaigatvensvlygnk---statckdkgnlksdgtagnnttagttgtvngfag 970
QY 1910 QTAVGASVGSAGAEIRIQNVAAAGEVSATSDAVNGSOLYKATQSIANATNELDHRHQH 1969
Db 971 atahgavsvgasgeerriqnvaageisatstdaingsqlvavakgvtn-----lagqvnkv 1026
QY 1970 ENKANAGISSMAMASMPQAVIPGRSMVTGGIATHNGQGVAVAVGLSKLSNDGQWVFING 2029
Db 1027 gkradagtasalaasqilpqasmpgksmvsiaagsyqgnglaigvrsisdngkviirslg 1086
QY 2030 SADTQGHVGAAGVAGGFHF 2047
Db 1087 ttnsqgktgvaagvggqww 1104

RESULT 10
AAB23859
ID AAB23859 standard; Protein; 1104 AA.
XX
AC AAB23859;
XX
DT 17-JAN-2001 (first entry)
XX
DE Haemophilus influenzae adhesin (Hia) protein from NTHi strain K22.
XX
KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
KW diagnosis; immunogenic; antigen.
XX
OS Haemophilus influenzae.
XX
PN WO200055191-A2.
XX
PD 21-SEP-2000.
XX
```

```
PF 16-MAR-2000; 2000WO-CA00289.
XX
PR 16-MAR-1999; 99US-0268347.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
WPI; 2000-618897/59.
XX
N-PSDB; AAA92498.
XX
PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating hemophilus influenzae
PT infection
XX
XX Claim 1; Fig 23; 275pp; English.
PS
XX The present sequence represents a Haemophilus influenzae adhesin (Hia)
XX protein from the non-typeable Haemophilus influenzae (NTHi) strain K22.
XX Hia genes and proteins have antiinflammatory, auditory and antibacterial
XX activities, and can be used in the production of a vaccine. An
XX immunogenic composition comprising an Hia gene, a polypeptide encoded
XX by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
XX protection against disease caused by Haemophilus strains in a
XX susceptible host, preferably a human. An Hia protein is useful as an
XX antigen, in immunogenic preparations including vaccines, as a carrier
XX for other immunogens, and in the generation of diagnostic reagents. Hia
XX is useful for treating diseases caused by the infection of Haemophilus
XX influenzae such as meningitis, epiglottitis, septicaemia and otitis
XX media. Recombinant production of Hia favours high recovery of the
XX protein compared to the low recovery of native protein from Haemophilus
XX influenzae species. A truncated protein has a significantly higher
XX amount of recovery than a full-length protein.
XX
SQ Sequence 1104 AA;
XX
Query Match 9.9%; Score 1016; DB 21; Length 1104;
Best Local Similarity 27.9%; Pred. No. 3,4e-37; Indels 394; Gaps 55;
Matches 373; Conservative 166; Mismatches 405;
QY 869 NKTSKVYDVNVDD-----TTIHLTGTDDNKKLGKVTTKLNKTS-ANGNTA----- 913
Db 2 nkifnvlwvvtwtwvvselttrahckcasatvavavatlalataaeannntsvtnglna 61
QY 914 ---TNFN-VNSDEDAVLNNAKDAENLTLAKEIHHTKTGTADTALQTFYTKKVDENNAD 969
Db 62 ygdtnfnttnnsiadlekvhqdaykgllnlnek-----dtnksflva----- 104
QY 970 DANAITVG-----QKNAN-----NQVNTLTLLKGENGLNI-KTDKNGTVTFGI 1010
Db 105 dntaatvgnlrklgwlskngtrneksyqvgadevltfgsgaatvsssskdgkhtiti 164
QY 1011 NTTSGLKAG-KSTLNDGGLSIKNPTGSEQIQVGADGVKFAKVNNGVGVGAGIDGTTTRITR 1069
Db 165 svtkgsfaevktdattgg-----qvnadrgkvkaeden---gadvd----- 202
QY 1070 DEIGFTGTNGSLDKSHPHLSKDGINAGCKKITNTIOSGEIAQNSHDAVTGCKIYDLKTELE 1129
Db 203 -----kkvatvk---dvakaındaatfvkvestdddie 232
QY 1130 NKISSTAKTAQNSLHEFSVADEQGNFTV---SNPYSSYDTSKTSVDVITFAGENGITTKV 1186
Db 233 ngaagknettdqal-----kagdtitlkagknkalkdngkksvtfalakdlidvlsak 285
QY 1187 NKGVVRVIGIDQTK-----GLTTPKLTGVNNGKGIIVIDSQNGQNTITGLSLTANVT 1238
Db 286 vsdklsigkdktnkvdiitsdanglklak--tgngng-----qngnvhlnlgiaatl---- 332
QY 1239 NDKGSVRTTEQGNLIKDEKTXRAASIVDVLISAGFNLOGNEAVDFVSTYDTVNFADGNAT 1298
Db 333 tdtitgmttqasngvavqnhnraasvadvinagwnlqngngasvdfvnaaydtvfvngtnt 392
```

```
QY 1299 TAKVYDDTSKTSKVYVD-----VNVDTTIEVK-----DKKLG--- 1332
Db 393 nvnvtcdtahkktvrvdvtgplpvgyvtedgkvtvkvdknyyvakqdgdsadmkkvenge 452
QY 1333 -VKTTTLSTGTGANKFALSQAOTDALVKASDIVA--HLNLTSLGDIQTAKGASQANNSA 1389
Db 453 laktkvklvsaagpnvklisnvaeg--teendavsfkqlkalq-ekqvlttasnayang 508
QY 1390 GYVDADGKNKI-----YSDTDNYYQAK--ND-----GTVDKTYKEVAKDKLV 1429
Db 509 gn-dadgkatgtlnnglnfkfstgdgellnlkventdvtftpkkgsv-qvgedgkatig 566
QY 1430 AQAOTPDCTLAQMNVKSVINKEQVNDANKKOGINEDNFAVFKLEKASDNKTKNAATVVG 1489
Db 567 ngtktdglveasel-----veslnklgwkvgvdkdgs-----g 600
QY 1490 DINAVAPLTFAGDTGTTAKKLGTELLIKGGQDTDTLNINIGVAVAGTDCGTVVKLAKD 1549
Db 601 eldgaasnetlvksgdkvll--kagenlkvk--qdgtl-----fyalakde 641
QY 1550 LTNLSNV-----NAGGTKIDDKGVSPVDSG-----QAKANTPVLSANGLDLGGKV 1595
Db 642 ltgvksvefkdtangsgastkktkdgittitsangangaatdadtikvvasdgisagnka 701
QY 1596 ISNVGKGTDKTDAANVQQLNEVRNLL-----GLNAGNDNADGNQVNIADIKKDPN 1646
Db 702 vknvsglkkfgdanfnpltsadnltkqyddaykgltnldekagdkltlvad-----n 756
QY 1647 SGSSNRVTRIKAGTVLGGKGNNDTEKLTATGGTQGVVDKDNANGDLSNV-WVKTKDKGSK 1705
Db 757 t-----aatv-----gdrlrglv-isaadktt 777
QY 1706 KALLATYNAAGOTNVLTNPAEADRIDNEQGRFRFHVNDGNOEPVVQGRNGIDSS---AS 1762
Db 778 gelnkeyna-----qvrnanevkfksq-nglhwsgkctvn 810
QY 1763 GKHSVAIGFOAKADGEAAVAIGRTQAGNOSTAIGDQAATGDQSIATGCTGNVYAGKHSQ 1822
Db 811 grrleit--felakdena-----lafgyskairdntvaigtgnvnaeksg 854
QY 1823 AIGDPSTV--KADNSYSVGNNOFTDATQDVFVGNN----- 1858
Db 855 afgdpnyiedkaggsyafgndnrit-skntfvlngvnaekykangdvdtetvdkdkgk 913
QY 1859 ---ITV-----TENSVALGSNSAISACTHAGTQAKKSDGTAGTATTAGATGVTKGRAG 1909
Db 914 ettvtvpkalgatvensvylgkn---statkdkgknksdgtagnatttagttgtvngfag 970
QY 1910 QTAVGAVSVGASGAEERRIQNVAAAGVSAATSDAVNGSOLYKATQSIANATNELDHRHQN 1969
Db 971 atahgavsvgsageerriqnvaageisatdangsglyavakgvtn-----lagqvknv 1026
QY 1970 ENKANAGISSAMAMSPQAYIPGRSMVTGGTATHNGQCAVAVGLSKLSDNGQWFKING 2029
Db 1027 gkradagtasalaasqlpqasmpkmsviagssyvggnglaigvrsidngkviirislq 1086
QY 2030 SADTQGHVGAAGAGCFHF 2047
Db 1087 ttnsggktgvaagvggyq 1104

RESULT 11
AAB23857
ID AAB23857 standard; Protein; 1004 AA.
XX
AC AAB23857;
XX
DT 17-JAN-2001 (first entry)
XX
DE Haemophilus influenzae adhesin (Hia) protein from NTHi strain M407.
XX
KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
non-typeable Haemophilus influenzae; antiinflammatory; auditory;
```

```
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
KW diagnosis; immunogenic; antigen.
OS Haemophilus influenzae.
PN WO200055191-A2.
PD 21-SEP-2000.
PF 16-MAR-2000; 2000WO-CA00289.
PR 16-MAR-1999; 99US-0268347.
PA (CONN-) CONNAUGHT LAB LTD.
PI Loosmore SM, Yang Y, Klein MH;
DR WPI; 2000-618897/59.
DR N-PSDB; AAA92496.
XX
```

Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for use as antigens and vaccines and for treating Hemophilus influenzae infection

Claim 1; Fig 21; 275pp; English.

The present sequence represents a Haemophilus influenzae adhesin (Hia) protein from the non-typeable Haemophilus influenzae (NTHi) strain M407. Hia genes and proteins have antiinflammatory, auditory and antibacterial activities, and can be used in the production of a vaccine. An immunogenic composition comprising an Hia gene, a polypeptide encoded by an Hia gene, or a recombinant Hia polypeptide is useful for inducing protection against disease caused by Haemophilus strains in a susceptible host, preferably a human. An Hia protein is useful as an antigen, in immunogenic preparations including vaccines, as a carrier for other immunogens, and in the generation of diagnostic reagents. Hia is useful for treating diseases caused by the infection of Haemophilus influenzae such as meningitis, epiglottitis, septicaemia and otitis media. Recombinant production of Hia favours high recovery of the protein compared to the low recovery of native protein from Haemophilus influenzae species. A truncated protein has a significantly higher amount of recovery than a full-length protein.

Sequence 1004 AA;

Query Match 9.4%; Score 970; DB 21; Length 1004;  
Best Local Similarity 28.1%; Pred No. 3.2e-35;  
Matches 341; Conservative 146; Mismatches 341; Indels 386; Gaps 44;

```
QY 979 KNANQVNTLTLLKGENGLNIKTDKNGTVTFGINTTSGLKAGKSTLMDGGLSIKNPTGSBQ 1038
Db 28 krasatvatavl-----atiltstvqattggttstnglkaygst----- 67
QY 1039 IQVGADGVKFAKVNNGVVGAGIDGTTTRTRDEIGFTGT-NGSLDKSKPHLSKDGINAGG 1097
Db 68 -----nnpnfnaagsatdlarq---fdgaydglmlnnekdanllvtdd 110
QY 1098 KKITNIQSGEITAQNSHDAVTGGKIYDLKTELENKISSATAQNSLHFESVADEQGNFT 1157
Db 111 k-----aatvnlrklglwlsksgtrnekssqqvkhadevifegkdgt 154
QY 1158 VSNPYSSYDTSKTSQD---VITFAGENGITTKVKNKGVVVRVIGIDQTKGLTTPKLTVGNN--- 1211
Db 155 v-----tsksengkhtvtftlekdlnvk-----natvsdklsilgangnk 193
QY 1212 -----NGKGIVIDSQGNQ---TITGLSNTLIANVTNDKGSVTRTEQGNIIKDEDKTRA 1261
Db 194 vdiatsdnglklfakpstngqngnhlgiastltdtit--gttksatng--vdvqgnhra 249
QY 1262 ASIVDVLGAGFNLCQNGEAVDFVSTYDTVNFADGNATTAKTAKTYDDTSKTSKVYDVND 1321
Db 250 asvadvlhagwnlqngasdvfntvdyfngvntnvnvt-tctahnkkt--vrvdv 306
```



Db	316		----	yv	317
Qy	1393	DADGNKVIYDSDNKNYYOAKNDCTVDKTKVEAKDKLV---	AQAQTPODQLAQMNVKSVIN	1449	
Db	318	tedsktvv--kvgnyeaekqgdsadmkkkvenglaktkvklysangt-npykisnva	d	374	
Qy	1450	KEQVND-----NKKGQINEDNAFVGLEKAASDNKTKNAATVVG-DLNAVAQTP	1498		
Db	375	gtedtdavafkqlalqdkqvklisanayang---gsdadggkatgtlgnlnfkfst	430		
Qy	1499	-----LTfAGDTGTtAKKLGET-----LTIKGGQDFTNKLTD-----NNIGWA	1537		
Db	431	dsellnikaaagtvtftpkgsyvqgdgkatigdgakttlglveaselvdslnkgkv	490		
Qy	1538	G-----TDG-----FTVKLAKDITJNLNSV---	1556		
Db	491	gvkgdgtgatdghtdtlvksghkvtlkagdnlkvkgegtnftyvlrdeitgvksvefd	550		
Qy	1557	-----NAGGTIDDKGVSPV-----DSSGOA--KANTPVLISANGLDGCKVISNYCKGPKDT	1606		
Db	551	tengangsakttkdgititpandangaaatdadikivasogdisagnkavknvvsnglkkf	610		
Qy	1607	DAANVOQLNEVRNLLJLGAGN-----NBADGNVNIAIDIKKDPSNGSSNRTRIVKAGTV	1661		
Db	611	gdanfnplt-----ssadnltkqyndaykgltcnleaksqkgtptvadnt--aatv	659		
Qy	1662	LGCIGNNDTEKLATGGIQGVGDKNANGDLSNV-WV--KTQDCSKKALLAIYNAAGQT	1718		
Db	660	-----gdrlrgwvisadkttgeskeysaqvrnanev	691		
Qy	1719	NYLTNPAEADRINEBOGRFFPHVDNGOEVPVOGRNGIDSASSGHSHVAIGFOAKADGE	1778		
Db	692	kfsngingivsgktldngrt-----eitfelakden	722		
Qy	1779	AAVAIGRQTQAGNQNSTAIGDNAQAQTGDOSIAIGTGNVVAGKHSGAIGDPSTV--KADNS	1836		
Db	723	a-----lafgskalrdntvaligtgnvnnaeksfagdnpolyledkagsy	768		
Qy	1837	SVGNNOFTDATOTDVFVGNN-----ITV-----TESN	1865		
Db	769	afgndnrhit-sknrtfvlgvwnaakyangdvdtetvkvdkdgettvtpkalgatven	827		
Qy	1866	SVALGSNSAISAGTHAGTOAKKSDGTAGTTTTAGATGTGVKGFAQTAVGAVSVGASGAER	1925		
Db	828	svylgnk----statkdkglnksdgtagnnttagttgtvngfagatahgahvsgvasgeer	884		
Qy	1926	RIONVAAGEVSATSITDAVNGSOLYKATOSTIANATNWELDRHTHONENKANAGISSAMAMAS	1985		
Db	885	rignvaeelsatstdalngslgylavakgvtn----lagvqnkvgrkradagtasalaasq	940		
Qy	1986	MPOAYIPGRSMVTGGIATHNGOGAVAGLSKLSNDGNQWVFVINGSADPTQGHVGAAGVAGCF	2045		
Db	941	lpqasmgskmsviagssvqqsgslai gyvrisldngkvilrslsgttnsqgktgvgaavgv	1000		
Qy	2046	HF 2047			
Db	1001	qw 1002			

RESULT	13	
AAW56319		
ID	AAW56319	standard; Protein; 2042 AA.
XX		
XX	AAW56319;	
XX		
DT	19-AUG-1998	(first entry)
XX		
DE	Haemophilus	paragallinarum antigenic protein #1.
XX		
KW	Haemophilus	paragallinarum; antigenic protein; HI antibody; diagnosis;
KW	vaccine;	chicken infectious coryza; CIC; fowl.
XX		
OS	Haemophilus	paragallinarum.

XX	Key	Location/Qualifiers	
XX	Peptide	1..70	
XX	Protein	/label= signal	
XX		71..2042	
XX		/note= "antigenic protein"	
XX	WO9812331-A1.		
XX	26-MAR-1998.		
XX	12-SEP-1997;	97WO-JP03222.	
XX	19-SEP-1996;	96JP-0271408.	
XX	(KAGA ) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.		
XX	(KAGA ) CHEMO-SERO-THERAPEUTIC RES INST.		
XX	Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;		
XX	WPI: 1998-230318/20.		
XX	N-PSDB; AAV22834.		
XX	Antigenic polypeptide from Haemophilus paragallinarum induces HI antibody production - and is useful for diagnosis of and preparation of vaccines for chicken infectious coryza		
XX	Claim 1; Page 51-69; 108pp; Japanese.		
XX	The present sequence represents an antigenic protein derived from Haemophilus paragallinarum strain A-221. The antigenic protein stimulates the production of HI antibodies in fowl. The protein and DNA coding for it can be used in the preparation of vaccines for the prevention of chicken infectious coryza (CIC). The protein and its antibodies can be used in the diagnosis and treatment of CIC		
XX	Sequence	2042 AA;	
XX		SQ	

[illegible]

Db 527 aladaqnaafalsqsfvessaatntitigvgsyakgnksflggtwastlsdrctvvlgnsts 586  
 QY 383 IKGAFTNALTNNICV---VKEADNSGLVKVLAKTLLNLTVEVNTTTLNATTVKVGSSS 439  
 Db 587 issqsg-nala---igvnnfigndaassialgmgti-----aksakspdsiaigkea 635  
 QY 440 STAEALLSLSLFTQPNTQSGST-----SKTVGVNGVKFTNNAETAAIG--TTRIT 490  
 Db 636 ridak-dtongtlygqvvydetrafrnfnessdyrmqamalgnakvrgsvgkmetgin 694  
 QY 491 RDKIGFARDGDVDEKOAPYLDKK-----OLKV-----GSVAI-----TIDNGIDAG 531  
 Db 695 smaigayacatlnsalgalvgsktdytwegletdpwvsegalsiptsgktgvisvgks 754  
 QY 532 NKISNLAKGSSANDAVTTEQLKAAP-----TLNAGAGISVTPTE-LSVDKAGSNVT 583  
 Db 755 errivnlasgsdtdavnvaqiktveerflseinnlqnggvkyisvektningdsgrva 814  
 QY 584 A-----PTYN--IGVKTTELNSDG-----TSDKFSVKSGSTNNSL-----VTAHLASLYLN 627  
 Db 815 sqirkgenyerykltqilyldarklngekfd-----qsnlkniravvqeleaeysg 868  
 QY 628 EVNRADSALQSTVKEEDDDANAITVAKDNTKNAGAVSILKLKCKNGLTVATKKDGTV 687  
 Db 869 elkttasalnqvatlqe-----vttmndkfnqyktqlen----- 905  
 QY 688 TFGLSQDSGLTICKSTYLNNDGLT-----VKDNEIOIQVANG-----IKF-----TNVNGSNP 735  
 Db 906 --asnadsarnvg-----glcpqatqikannnyindgakqgdsiafgwqektesgan- 955  
 QY 736 GTGIANARITRDKIGPAGSDGAVTNKPYLDQDKLVQV-NVKITNTGINAGKAITGLS 794  
 Db 956 -nlagkqala--igfqansa-----enaisigtntsmtsgavaigkatvta 1002  
 QY 795 PTLPSIQDSQSRNLEGNTTQDKDKSNAASINDLLTGFNKLNNNPIDFVSYDIDVEA 854  
 Db 1003 gqkpsia-----lg--qdstvansa-----istspnm--inglifnfa 1038  
 QY 855 NGNATTATVTHDTANKTSKVYDVNVDDTTHLTGTDDNKKLGKVTTKLNKT--SANGNT 912  
 Db 1039 gspetlgvisigtagrerkiv-nvaagdvsgasteaingsqlyatnfmiskvagsksnf 1097  
 QY 913 ATNFVNSSDEDLVNAKDAENLNTLAKELHTTKGTADTALQTFVK-----KVDENN-- 966  
 Db 1098 ggnvnlg--dgtitftnig-----gtqatihdaiannvltkgylkadgnpdt 1144  
 QY 967 ----NADDANAITVQGNK--ANNQVNTLTLKGENGNIKTDKNGTTFGTGINTTGLK-- 1017  
 Db 1145 gnqgqkvvelgnalilsatnqannvnykt-----nrltynsqngtilfgmredpsvkqi 1200  
 QY 1018 -AGKSTLNDGSLSTKNPTGSEIQVAGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFTG 1076  
 Db 1201 tag-----tynttgd-----annknqinntlqgt-----tleatgits 1233  
 QY 1077 TNGSLDKSKPHLSKXGI--NAGGKTIQSGEITAQNSHDAVGTGGKIYDLKTEL--ENKI 1132  
 Db 1234 svgsstnyagfslgadsvtfskggagtvkissvgsdatdtdaatlkqvkkeyrttlvgdndi 1293  
 QY 1133 SSTAK---TAQNSLHEFSVADEQGNFTVSNPYSYDTSKTSDDVITTFAGENGITVKNKG 1189  
 Db 1294 taadrsqtsngitynlsl-----nkgvtsateekvsvgkt---vyealnait-----gn 1341  
 QY 1190 VYRVGIDQTKGLTTPKPLVGNNGKGIVIDSQNGQNTITGLSNTLIANVTNDKGSVRTTEQ 1249  
 Db 1342 iftigldtd---tlnk--innpadqdlslsesgknaitg----lvdvvktnspitvpep 1392  
 QY 1250 GNIIDEKDKTRAASIVDVLNAGFNLQNGEAVDFVSTYDTVNFADGNATTAKVYDIDTSK 1309  
 Db 1393 st---dsnkktftv-----gvdfdt-----itegdtdckklttskv 1429  
 QY 1310 TSKVYVDVNVDDTTTEVKDKKLGKVTTLTSTGTGANKFA-----LSNQATGALVKAS 1363  
 Db 1430 esyvtnkianfstlildsdgrsgnatt--andvgkrrlsdgtfiksensftlsgkqyns 1487

QY 1364 DIVAHLNLTSGDI-----QTAKGASQANNAGSYVDADGNKNKIYDSTDNKKYQAKNDGTV 1417  
 Db 1488 dslgmyddqngvfklsimlmtalttslant---fakldasnltddsnkekwrtaln--vy 1542  
 QY 1418 DTKT---EVAKDKLVAQAOTPDGTLAQMNVKSYINKQOVNDANKKOGINEDNAFVKGLEK 1474  
 Db 1543 sktevdaelqskvltl---tpdegl-----ifatkagsgn-nagldagnkksisnv-- 1589  
 QY 1475 AASD-NKTKNAAVTVGDILNAVAQTPPLTFAGDTGTAKKLGETLTIKGGOTDTINKLTDNNI 1533  
 Db 1590 adgdisptsgdvtrgrqlyalmkgirvygd-----evaptkqtaptanap-tatta 1642  
 QY 1534 GVAAGTGTFTVKLAKDLTNLNSVNAVAGTKIDDKGVSVFSDSSGQAKANTPVLSANGLD--- 1590  
 Db 1643 ptasstgwa-----ttant--agg--vapagnvatgdiaptqptlpemntalvdohl 1691  
 QY 1591 ---LGGKVISNVGKTGTDTAANVQ-----OLN-EVRNLLGLGNAGNDNA-----DGNQV 1636  
 Db 1692 avp199sliklbgdhvktlisdahqvgislpnlsienlviigandpekeklaaeqnal 1751  
 QY 1637 NIADIKKDPNS-----GSSSNRTVI-----KAGTVLGGK-----GNNDTEKLTATGGIOV 1680  
 Db 1752 vitn-kddgnaamvfnnekmlvlsdkeakprvilldgngaltivgnddeq--vtlsakk 1808  
 QY 1681 GVDKDGANGDLNVVWVKTQKDGSKALLATYNAAGOTNYLTNPAAEIDRINEQGIREF 1740  
 Db 1809 gkdidgn---dlrslsvttet-----nadgglekvetsfatmdd-----gikf- 1849  
 QY 1741 HVNDGNQEPVVOGRNGIDSSASGKHSVAIGFAKADGPAVAIGRQTOAGNQSTAIAGDNA 1800  
 Db 1850 -----kagdkvl-----nkklnetv 1865  
 QY 1801 QATGQOSTAIGTCNVVAGKHSAGIDPSTVVKADNSYSGVNNNOFTDATQTDVFGVGNNT 1860  
 Db 1866 elvgdevnt-----tsitddnkvvkslnkklia-----idevkiptndp 1903  
 QY 1861 VTESNSVALGNSAITSAGTHAGTQAKSDGTAGTTTAGATCTVKGFAGTAVGAVSVGA 1920  
 Db 1904 daqgdsivinnngghagnkvitgvkasd----- 1932  
 QY 1921 SGERRIONVAAGEVSATSDAVNGSQLYKATQSTANATNELDRIHONENKANAGISSA 1980  
 Db 1933 -----dptsavnrgqlntvldnvgnfnqvnqrldltresraglaga 1975  
 QY 1981 MAMASMPQAVIPGRSMVTGCIATHNGOGAVAVGLSKLSDNGQWVFKINGSADTQGHVGA 2040  
 Db 1976 matasiqnvalpqtktisvgtatfkgenavaigmrlsdngkvirlsgmstsgndkga 2035  
 QY 2041 VGAGFHF 2047  
 Db 2036 msvgfsf 2042  
 RESULT 14  
 AAW56322  
 ID AAW56322 standard; Protein; 2039 AA.  
 XX  
 AC AAW56322;  
 XX  
 DT 19-AUG-1998 (first entry)  
 XX  
 DE Haemophilus paragallinarum antigenic protein #2.  
 XX  
 KW Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;  
 KW vaccine; chicken infectious coryza; CIC; fowl.  
 XX  
 OS Haemophilus paragallinarum.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..70  
 FT /label= signal  
 FT Protein 71..2039

```

FT      /note= "antigenic protein"
PN      W09812331-AL.
PD      26-MAR-1998.
PF      12-SEP-1997; 97WO-JP03222.
PX      19-SEP-1996; 96JP-0271408.
PA      (KAGA ) 2H KAGAKU & KESSEI RYOHO KENKYUSHO.
PA      (KAGA ) CEMO-SERO-THERAPEUTIC RES INST.
PI      Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;
XX      WPI; 1998-230318/20.
DR      N-PSDB; AAV22837.
XX      Antigenic polypeptide from Haemophilus paragonillarum induces HI
PT      antibody production - and is useful for diagnosis of and preparation
PT      of vaccines for chicken infectious coryza
XX      Claim 5; Page 71-87; 108pp; Japanese.
CC      The present sequence represents an antigenic protein derived from
CC      Haemophilus paragonillarum strain C-53-47. The antigenic protein
CC      stimulates the production of HI antibodies in fowl. The protein
CC      and DNA coding for it can be used in the preparation of vaccines
CC      for the prevention of chicken infectious coryza (CIC). The protein
CC      and its antibodies can be used in the diagnosis and treatment of CIC.
XX      Sequence 2039 AA;
SQ

```

Query Match 7.8%; Score 804; DB 19; Length 2039;  
Best Local Similarity 21.8%; Pred. No. 1.5e-27;  
Matches 514; Conservative 298; Mismatches 778; Indels 772; Gaps 103;

```

QY 12 ATGTMAVAEYAKSHSTGGGSCATGQGVSVCTLSFARIAALAVLVIGATLGSAYAKRDK 71
DB 124 atgs-iaigqakndwatsrlalggaknesiasdval-----snainrf 169
QY 72 TKHIAIG---EONPRSGTAKADGDRAIAGENANAGGQAIAGSS--NKTVNGSSL 125
DB 170 kksivigIntyqlprrapesr---qgsvigenaksgnqsvslgnawskt----- 220
QY 126 DRIGTDATQESIAIGDVKASGDASIAIGSDLLHLLDQGNPKHPKGT----- 174
DB 221 -----nsisigagtfaegkstiaigt--kilgtynndkjpapswdgrtgkaptn 268
QY 175 ----LINDLINGHAVLKEIRSKON-----DVRYR----- 200
DB 269 siwdifselymgkktngtdydkndrdpnkpeafytsdfksryvnnpstpsptyaaklg 328
QY 201 -----RTASCHASTAVGANSYAGCHESNAGFTRA-TAKSAY-SLAVGLAATAGCQSTI 252
DB 329 aialgsrtiaagemstavgslafaldkstanglrsvfakdavggtalgeesrtfakds 388
QY 253 AIGSDATSSLSAIALGAGTRAOLGCSIALG--QGSVWTSQDNN-----SRPAYTPNTQ 304
DB 389 aignkteesnagsmgygkavagagaiaigaevaageafdsqagnlllirgayatiks 448
QY 305 ALDPKFEQATNNTKAGPLSIGNSIKRKTIINVAGVKNKTDVNVVAOLEAVVAKWAKERRITF 364
DB 449 a-----dkssddikag-----dainvftqffdmnltqgshly 480
QY 365 QGDDNSTDVKGLDNLTIKGAETNALTNNIGVVKKADNSGL-KVKLAKTLNMLT--- 420
DB 481 enttyltsagdkiklaavdggnknaignkktfaskansvalgsyalasaqafalgs 540
QY 421 -----EVNT-----TTLNATTYKVGSSSTTAELLSDSLTF 452
DB 541 yslvsplaantivigvggyatgsnsvfsgsvwstlsartvv-lgysasissd-shdsiam 598

```

```

QY 453 -TQPNITGSGQSTSKTYGV-----NGVKFTNNA--ETT 481
DB 599 gvnafingnsnslalgtgstiaknakspdsiaigkdsridakldngvlytpqvydett 658
QY 482 AAIGT-----TRITRDK-----IGFARDGDVDEKQA----- 507
DB 659 raftrfdenkymrdamalgnaksvrgkgmetginsmaigarsqatignatvgnak 718
QY 508 -----PYLDKKOLKV---GSVAITIDNGIDAGNKKISLNIAKSGSANDAVTEOL 553
DB 719 tdytwegleadpwwskgaistsptsgkigv-isvgsgkserriinvvasglddavnvaql 777
QY 554 KAAK-----PTLNAGAGISVTPTE-ISVDKASGNVTA-----PTYN--IGVKTTELNS 598
DB 778 ktieerfqseidlqngggvqylsvektningeagrvasqickgesykyvklktqllyl 837
QY 599 DG-----TSDFKFSVKGSGTNNSLV---TAEH---LASYLENVAHPADSALQSTFV----- 642
DB 838 darklungekfdqtsldkiskavqeleacysgelkttaselnrvamlnaettvndfgkf 897
QY 643 ---KEEDDDANAITVAKDTTKNAGAVS---ILKLKGNG-LTVATKKDGTVTFGLSQDS 695
DB 898 ngyktqienatna-----dseknvgglspqvaiakannnyindgakgdsiafwgq--- 949
QY 696 GLTIGKSTLNDGLJVYKDTNEQIQVGANGIKFTNVNGSNPGTGIANTARITRDKIGFAGS 755
DB 950 ----aktseannglagkq-----ala-----igfqan 972
QY 756 DGAVDTNPYLDQDKLQVG-NVKITNTGINAGGKAITGLSLPSIADSSRNIELGNTY 814
DB 973 ssa-----enaisgtndtsmtgavaigkatvtagdkpsia-----lg--- 1012
QY 815 QDKKSNA--SINDILNTGFNLKNNNPFIDFVSTYDIVDFANGNATTATVTHDTAKNTS 872
DB 1013 qdstvansaisrtssvmingltfn-----fagspetlgvisigtagker 1057
QY 873 KVVYDVNVDDTTIHLTGTD--DNKKILGVKTKLNKTS-----ANGNTATN----- 915
DB 1058 kl--vnvaagdisqstsealngsqlyatnfmnlkagvktknfgnpanlatdgtitfcn 1114
QY 916 -----FNVNSDEDALYNADIAENLNTLAKEIHTTGTGTADTALQ 955
DB 1115 iggtgqgdthdainnvltklislstaeveevsgeavydalkgkaptvsaeankgitgl- 1173
QY 956 TPTVKKVDENNADANAITVQGNKANNNOVNTLT---LKGENGLN-IKTDKNG---TVT 1007
DB 1174 vdvvkan-----spitvepstdnkktkftvglmkiegvnhtfdksgqdlngvlt 1225
QY 1008 FGINTTSLGLKAGK-----STLNDGGLSIRKPTGSEQ-----IQVGADGVK 1047
DB 1226 -grmsagltfkkgdtngstttfaedgltidstnsaqlnvkvsrdgfsvkngsdeak 1284
QY 1048 FAKVNNNGVVGAGIDGTRITRDEIGFTGTNGSLDKSRPHLSKDGINAGCK-----KIT 1101
DB 1285 lastkis--igaenaehvevtksialkadnts-dkssitlaqdaaitlagntatgtaikt 1341
QY 1102 NIQSEIAQNSHDVATGGKIYDL-----KTELENKISSTAKTAONSLHEFSV 1148
DB 1342 gvadgnitvnskdavngggqirtllgydsgakiggtekttsiseaisdvkdaldlayka 1401
QY 1149 ADEQGNNEFTVSNPYSDTSKTSIDTIFAGENGITTKV-NKGVVRVGI-DOTKGLTTPKL 1206
DB 1402 dnknkg-----tvkltdglnfstnidasvedngvkvftlkdkltgikt--- 1446
QY 1207 TVGNNGKGIIVIDSON-QONTITGLSNTLIANTYNDKGSVRT--TEOGNIITKDEKTRAASI 1264
DB 1447 -----iatesinasqniiaaggtvtvgtet--egivltksgsgn-----drtls--- 1487
QY 1265 VDVLGAGNLONGEAVDFVSTYDVNFADGNATTAKTYTDDTSKTSKVYDVNVDDTTI 1324
DB 1488 ---lsgagnaatdgiikvgvka-----gtadtdavnkgqldklfkainda----- 1529

```



QY 1325 EVKDKKLGKVTTLTSTGTCGANKFALSNOAQTDGDLVKKASOTIVAHNLTLSDGIQPAKASQ 1384  
 Db 1530 -----lg--tttlavt-knpnqntifnplngtapttkfkdavdkltta-----vntgwsgskv 1577  
 QY 1385 ANNSAGY--VDADGNKVIYDSTDNKYYQAKNDGTVDKTKEVAKDKLVAQAQTPDGTILAQM 1442  
 Db 1578 gilatgidida-gnkkisnvadgdisptsgd----- 1608  
 QY 1443 NVKSVINKEQVNDANKQGI---NEDNAFVKGLEKAAADNKTNAAVTVGDLNAVAQTPL 1499  
 Db 1609 -----vvtgrql-almqkgirvygdevspkttaptasataggattantagvavapagn 1663  
 QY 1500 TFAGDTGTTAKKIGETLITIKGGQDTNKLVDNIGVAGTGTFTVKLAKDLTLNLSNVAG 1559  
 Db 1664 vatgdaptgpalpemkt-----alvgdhlaplpg---slkdhghnvkttisag 1711  
 QY 1560 GTRIDDKGVSF-----VDSQOAKANTPVLISANGLDLGGKVISNGVGTDKDIDA 1608  
 Db 1712 -----nqvqislpnoisniennlvigsnkpekaklaaegnal-----vitn-----kddgna 1758  
 QY 1509 ANVOQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNS---GSSSNRTVIRKAGTVLGGK 1665  
 Db 1759 amv--fnnekmlv1-----sdkkakpravldgngaltiv----- 1792  
 QY 1666 GNDTETKATGGIQGVKDGNGDLSNVVKTQDGSKKALLATYNAAQOTNYLTNNP 1725  
 Db 1793 gnddsq--vtlsakkgkldgn---disrlsvttert-----nadgqiekvetsf 1837  
 QY 1726 AEADRINEOIGIRFFHVDNQNQEPVQVGRNGIDSSAGKHSVAIGFOAKADGEAAVAIGR 1785  
 Db 1838 atmdd-----glkf-----kdgdkvi----- 1854  
 QY 1786 QTAGNQSIAGNNAQTGQSTAIQTGNVAGKHSAGIDPSTVRADNSYSYGVNNNQFT 1845  
 Db 1855 -----nkklnetvelvgdenvt-----tsltdnkvkslnkklia 1889  
 QY 1846 DATQTDVFGVGNITVTESNVAGLSAISAGTHAGTQAKKSDGTAGTTTATGATGTVK 1905  
 Db 1890 -----idevklpntdpdaqgdsivlmgghagnkvitgvksad----- 1929  
 QY 1906 GFAGOTAVGAVSVGASGAERRIONVAAGEVSATSDAVNGSOLYKATQSIANATNELDHR 1965  
 Db 1930 -----dptsavnrqgintvldnvgqnfqngqr 1957  
 QY 1966 IHQENKANAGISSAMAMSPQAYTPGRSMVTGGIATHNGQAVAGVLSKLSNDGQWVF 2025  
 Db 1958 igdltresragiagamatalsqvalpgkttisvgtatfkgenavaigmrlsngdkvgi 2017  
 QY 2026 KINGSADTQGHVGAAGVAGGFHF 2047  
 Db 2018 rlsgmatngdkgaamsvgftf 2039

## RESULT 15

AAB23858  
 ID AAB23858 standard; Protein; 1094 AA.

XX AC AAB23858;

XX DT 17-JAN-2001 (first entry)

XX DE Haemophilus influenzae adhesin (Hia) protein from NTHi strain K9.

XX KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;  
 KW non-typeable Haemophilus influenzae; antinflammatory; auditory;  
 KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;  
 KW diagnosis; immunogenic; antigen.

XX OS Haemophilus influenzae.

XX PN WO200055191-A2.

XX XX 21-SEP-2000.

PD

XX

PF 16-MAR-2000; 2000WO-CA00289.

XX

PR 16-MAR-1999; 99US-0268347.

XX

PA (CONN-) CONNAUGHT LAB LTD.

XX

PI Loosmore SM, Yang Y, Klein MH;

XX

XX WPI: 2000-618897/59.

DR

DR N-PSDB; AAN92497.

XX

PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for  
 PT use as antigens and vaccines and for treating Hemophilus influenzae  
 PT infection

XX

PS Claim 1; Fig 22; 275pp; English.

XX

CC The present sequence represents a Haemophilus influenzae adhesin (Hia)  
 CC protein from the non-typeable Haemophilus influenzae (NTHi) strain K9.  
 CC Hia genes and proteins have antinflammatory, auditory and antibacterial  
 CC activities, and can be used in the production of a vaccine. An  
 CC immunogenic composition comprising an Hia gene, a polypeptide encoded  
 CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
 CC protection against disease caused by Haemophilus strains in a  
 CC susceptible host, preferably a human. An Hia protein is useful as an  
 CC antigen, in immunogenic preparations including vaccines, as a carrier  
 CC for other immunogens, and in the generation of diagnostic reagents. Hia  
 CC is useful for treating diseases caused by the infection of Haemophilus  
 CC influenzae such as meningitis, epiglottitis, septicaemia and otitis  
 CC media. Recombinant production of Hia favours high recovery of the  
 CC protein compared to the low recovery of native protein from Haemophilus  
 CC influenzae species. A truncated protein has a significantly higher  
 CC amount of recovery than a full-length protein.

XX Sequence 1094 AA;

Query Match

Best Local Similarity 6.8%; Score 697.5; DB 21; Length 1094;

Matches 316; Conservative 161; Mismatches 416; Indels 387; Gaps 58;

QY 944 HTTKGTADTA---LQFTVKKVDENNADANAIV-GOKN-----ANNQVNTLTUK--- 991

Db 26 htkrasatavatavlatqlsataeanssavtsrlvnyvgdntkfnnaansiadlnkqndg 85

QY 992 -----GENGLNIK--TDKNGTVTFG-----INTTSLKACKSTLNDGGLSTKNP 1033

Db 86 vhdglllnlengankllvddntaatvgdlrklgwvvstkng-----ken 130

QY 1034 TGSEQIQVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFTGTNGSLDKSPHLSKDG 1093

Db 131 eksqgvk-----qadevlfksgk----- 149

QY 1094 NAGCKKITNIQSIEIAQNSHDVATGCKIYDLKTELENKISSAKTAQNSLHEFSVADEQG 1153

Db 150 -----gvqvtstseengkhaite---falakldmr----- 175

QY 1154 NNFTSNPYSYDTSKTSIDVITFAGENGITTVKNKGVWRVGDQTKGLTTPKLTIV----- 1208

Db 176 -----tatvsdtlitggs---tt-----tgsattpkvvnvctas 206

QY 1209 GNNNGKVIDSONGQNTI--TGLSNTLANVNDKGSVRTTEQGNITIKDEKTRAASIVD 1266

Db 207 glnfakgat--gangdttvhltniaatlqdtllntqvvskld--ngitadekkrasavqd 263

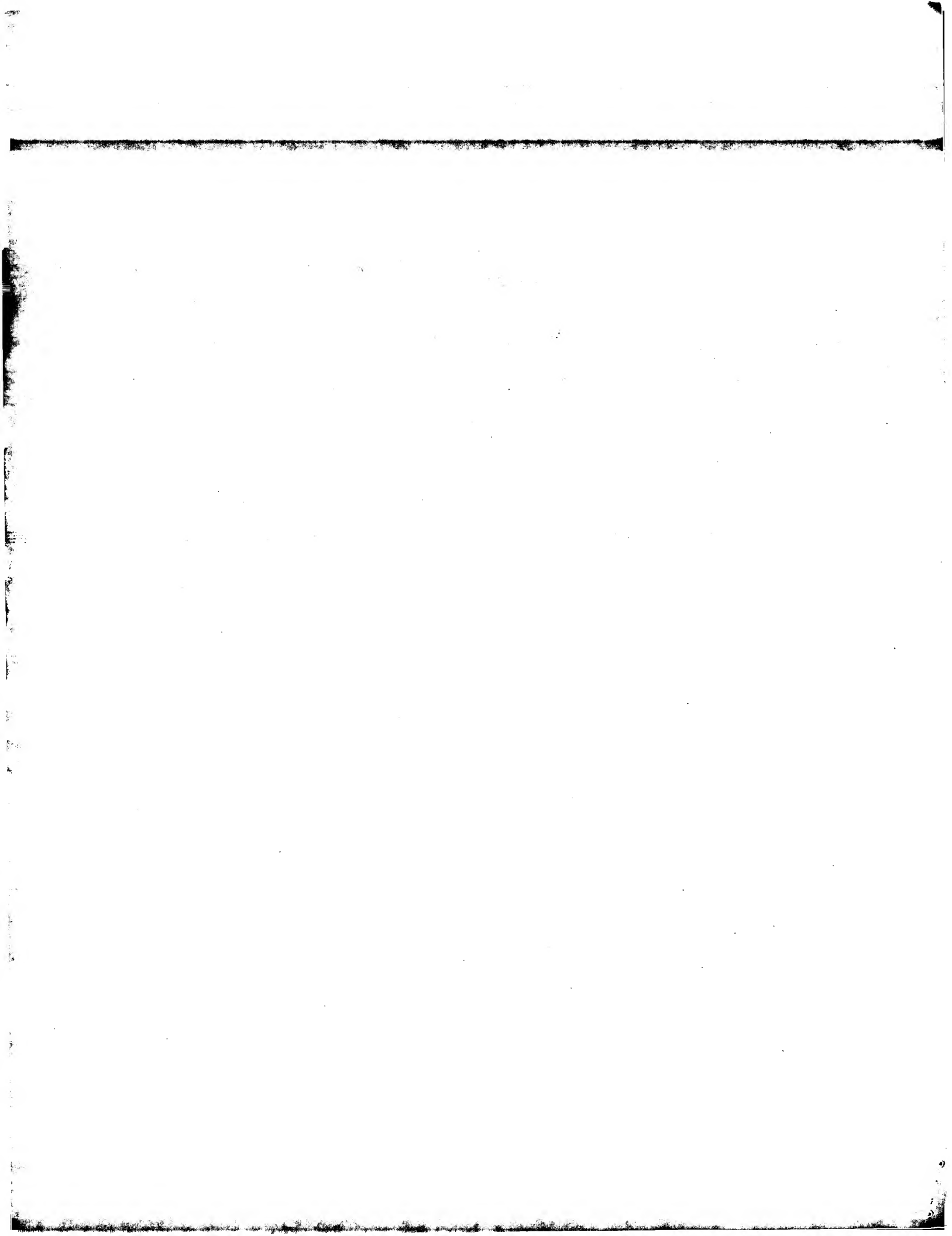
QY 1267 VLSAGFNLOG-----NGEAVDFVSTYDTVNFADGNATTAKTVDYDTSKTSKVYDVNDV 1320

Db 264 vlnsgwnlkgvktgattsdndvfrtydtevfisgeettllvtvdsesngsktckvkgak 323

QY 1321 DTTIEVKKKGLGKVTTLTSTGTG---ANKFALSNOA---TGDALVKASDVAHLNLTLS 1373

Db 324 tsvikekdgkl-----ftgkankdntqvasnaaddddegklvtaetvlnavnkag 375

Qy	1374	GDlQTAKGASQANNsAGYVD-----ADGN-----KVlYDSTDN-----KYyQAKNDG--	1411
Db	376	wrIkT-----tgnandggfetytsgtntvfdagngttlavvtgdatnglttkykeavgdgl	431
Qy	1416	-----TVDKTK-EVAKDKLVAQAQTPDGTCLAOMNVKSVINKEQVNDANKKOGINEDN	1466
Db	432	kIgdqkIcadttaltvtggkvt-----pdatt-----pgklvlnasgladalnK-----	476
Qy	1467	AFVKGLEKASDNKTKNAAVTVGDLNVAQAQTPLTfPAGDTGTAKKLGTTLITIKGGQTDTN	1526
Db	477	--lswtakaead--tang-----geId-----gtadekevkagetvtfkagK-----	514
Qy	1527	KLTDNlGvVAGTGDGTfVKLAKDLNLNSV-----NAGGTKlDDKGVSFVDSsGQA--K	1578
Db	515	-----nlkvqdgantfyslgaltgltstlgtgnngaktelnkdglttitpangagann	569
Qy	1579	ANTPVLsANGLDLGGKVISNVGKGTfKDTDAANVOOLNEVRNLL-----GLGNAGND	1629
Db	570	antIsVtkdgIsagqgsvknvvsglkkfgdanfdplttessadnltkydyddaykgltlnldek	629
Qy	1630	NADGNQVNlAD-----IKKDPNSG---SSSnrTfVlKAGTV--LGKGKND	1669
Db	630	gadkqtltvadntaatavglrlglvwIsadkttgldkeynaqvrnanevkfksgnginv	689
Qy	1670	TEKLTAGTGGVGVDRDGNAGDL--SNVvVYTKRDGSKKALLAI--YNAAQGTNYLTNN	1724
Db	690	sgktvngrrreItfel-----akgevksneftvketngktsIvkgdkyyskedIdpatgk	746
Qy	1725	PAEAIDRIEQGIRfFHVNDGNOEPV-VQGRNGIDSSASGKHSVAIGFQAKADGEAAVAI	1783
Db	747	p-----kvtnngnavaakygdkgdkgvvsadgsntavltlnkgyyv-----	787
Qy	1784	GROTQAGNOSlAGDNAAQTGPOOSTAIGTGNVVAKHSGATGDPS-----TVKADN	1834
Db	788	-----lgnqv-----adaIaksgfelglad--aekakafgdetkalsadkIetvnand	834
Qy	1835	S--YsVGNKNQfTATOTDFVGvGNlNlVt--ESNSVALGSNSATsAGThAGTQAK-----	1886
Db	835	kvrfaIngltkvsaaIvesIdandgdkvtftfktdvelpltcqIyntdangkkIvknngdkw	894
Qy	1887	---KSDGTAGTt---TtAGATGT-----VK--GFAGQTAV-----G	1914
Db	895	yytkdgdsgtdmktvltlgnvdsdgkvvvkednkvyhvksgdsgtdktqtvveeakvstdekH	954
Qy	1915	AVsVG-----ASGAERRlQNVAAGEVSATsTDAVNGSOLY----KATOSlANATNlELDHRIH	1967
Db	955	vvsIdpndgskgkvvInnmangeIsatstdaIngslIyvakgvtnlagqvvnnlegkvn	1014
Qy	1968	QNEKNKANlSSAMASMPQAYlPGRSvWVTGGTlATHNGQCAVAVGLSKLSDNGOWVFKI	2027
Db	1015	kvqkradagtasalaasqlpqetmpgksmvslagsvygqnglaIgvrrIsIdngkvIrl	1074
Qy	2028	NGSADTQGHVGAAGVAGfHF	2047
Db	1075	sattnsqaKtgvaagvgvqaw	1094



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2001, 23:14:44 ; Search time 44.72 Seconds  
(without alignments)  
942.496 Million cell updates/sec

Title: US-09-361-619-7  
Perfect score: 10303  
Sequence: 1 MNHIYKVFNKATGTFMAVA.....NGSADTQGHVGAAGAGPHF 2047

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/pCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7380	71.6	2123	4	US-08-968-685A-10
2	1123	10.9	2353	4	US-08-377-155-33
3	1123	10.9	2353	4	US-08-913-942-4
4	792.5	7.7	1912	1	US-08-409-995-4
5	792.5	7.7	1912	3	US-08-685-467-4
6	678.5	6.6	1098	1	US-08-409-995-2
7	678.5	6.6	1098	3	US-08-685-467-2
8	678.5	6.6	1098	4	US-08-377-155-32
9	678.5	6.6	1098	4	US-08-913-942-2
10	573	5.6	679	4	US-08-913-942-15
11	517.5	5.0	1600	2	US-08-617-697-10
12	503.5	4.9	1599	2	US-08-617-697-9
13	501.5	4.9	1529	2	US-08-728-470-10
14	501.5	4.9	1529	4	US-08-719-641-10
15	457	4.4	1338	2	US-08-728-470-9
16	457	4.4	1338	4	US-08-719-641-9
17	435	4.2	1612	1	US-08-169-927-2
18	433	4.2	1536	1	US-08-038-682-2
19	433	4.2	1536	1	US-08-302-832-2
20	433	4.2	1536	2	US-08-530-198-2
21	433	4.2	1536	2	US-08-469-880-2
22	433	4.2	1536	2	US-08-728-470-2
23	433	4.2	1536	2	US-08-617-697-2
24	433	4.2	1536	4	US-08-719-641-2
25	418.5	4.1	1477	1	US-08-038-682-4
26	418.5	4.1	1477	1	US-08-302-832-4
27	418.5	4.1	1477	2	US-08-530-198-4

28	418.5	4.1	1477	2	US-08-469-880-4	Sequence 4, Appli
29	418.5	4.1	1477	2	US-08-728-470-4	Sequence 4, Appli
30	418.5	4.1	1477	2	US-08-617-697-4	Sequence 4, Appli
31	418.5	4.1	1477	4	US-08-719-641-4	Sequence 4, Appli
32	398.5	3.9	599	4	US-09-377-155-15	Sequence 15, Appli
33	393.5	3.8	594	4	US-09-377-155-9	Sequence 9, Appli
34	392.5	3.8	598	4	US-09-377-155-5	Sequence 5, Appli
35	391.5	3.8	598	4	US-09-377-155-13	Sequence 13, Appli
36	384	3.7	589	4	US-09-377-155-19	Sequence 19, Appli
37	383.5	3.7	592	4	US-09-377-155-17	Sequence 17, Appli
38	383.5	3.7	594	4	US-09-377-155-7	Sequence 7, Appli
39	378	3.7	592	4	US-09-377-155-2	Sequence 2, Appli
40	375	3.6	591	4	US-09-377-155-11	Sequence 11, Appli
41	375	3.6	591	4	US-09-377-155-21	Sequence 21, Appli
42	368.5	3.6	658	1	US-08-409-995-5	Sequence 5, Appli
43	368.5	3.6	658	3	US-08-685-467-5	Sequence 5, Appli
44	368.5	3.6	658	4	US-08-913-942-5	Sequence 5, Appli
45	343.5	3.3	607	1	US-08-409-995-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-08-968-685A-10  
; Sequence 10, Application US/08968685A  
; Patent No. 6214981  
; GENERAL INFORMATION:  
; APPLICANT: TUCKER, KENNETH  
; APPLICANT: PLOSTLA, LAURA  
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE  
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/968 685A  
; FILING DATE: No. 6214981ember 12, 1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baldwin, Geraldine F.  
; REGISTRATION NUMBER: 31,232  
; REFERENCE/DOCKET NUMBER: 7969-060  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2123 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-968-685A-10  
  
Query Match 71.6%; Score 7380; DB 4; Length 2123;  
Best Local Similarity 68.3%; Pred. No. 0;  
Matches 1555; Conservative 118; Mismatches 219; Indels 386; Gaps 20;  
  
QY 1 MNHIYKVFNKATGTFMAVAEYAKSHSTGGSCATGCGVCTLSFARIAALAVIGAT 60  
|||||

Db 1 MNHKKVFNKATGTFMAVAEYAKSHSTGGSCATQGVGSRVTLSPFARIAALAVLIGAT 60  
 Qy 61 LSGSAYAKKOTKHIAIGBQONPRSGCTAKADGDRAIAIGENANAGGGAIAIGSSNKT 120  
 Db 61 LNCSSAYAKKOTKHIAIGBQONPRSGCTAKADGDRAIAIGENANAGGGAIAIGSSNKT 112  
 Qy 121 NGSSLDKIGTADGOSIAIGDVGKASGASIAIGSDDLHLDDQHGNPKHKGKTLINDL- 179  
 Db 113 NSNNANICAKASGNSIAIGDVGKASGASIAIGSDDLHLDDQHGNPKHKGKTLINDL- 165  
 Qy 180 --INGHAVLKEIRSSKNDV-KYRRTTASGHASTAVGAMSYAOGHFSNAGFTRAKSAY 236  
 Db 166 PIIRGOKALNDIYQLADTLNQRTHAQGHASTAVGAMSYAOGHFSNAGFTRAKSAY 225  
 Qy 237 SLAVGLAATAGOSTAIGSDATSSSIALGAGCRAOGLSIALGOGSVVYQSDNNSR 296  
 Db 226 SLAVGLATATAKAASSIAGSAGSIAIGDVGKASGASIAIGSDDLHLDDQHGNPKHKGKTLINDL- 285  
 Qy 297 P----AYTP-NTQALDPKFKATNNTRKAGPL-----SIGNSIKRKTIINYGAGVKNKTDAYN 346  
 Db 286 AANVRAYAPDNDQPIDNRYKATFKNGATDVESIGNSNGNDSIRKLIINVAGSADTDAYN 345  
 Qy 347 VAOLEAVVWAKERRITFOGDNSTOVKICLONTLTIKGAETNALTNNIGVWKEADNS 406  
 Db 346 VAOLKEAVRLA-NRQITFKGDSNNRVEKGLKTLITITGGAOTSAITDHNIGVWQNG-- 402  
 Qy 407 GLKVKLAKTLNLTTEVNTTTLNATTTVKVGSSSTTAEILLSLSTFTOPTGTSQSTKTV 466  
 Db 403 GLKVLAEILTLKMWTTENLTANEKVTVGK----- 433  
 Qy 467 YGVNGVKTNNATTAAGTTRIRDRKIGFARDGVDEKQAPYLDKQKLVKGSVAITON 526  
 Db 434 -----TRUTDKIGFTNDMNGIDSKPYLDK-----DT 461  
 Qy 527 GIDAGNKKISNLAKGSANDAVIEOLKAAPTLNAGAGSIVTPTTEISVDKSGNVTAPT 586  
 Db 462 GIHAGGOKTKLTVAGVDDDAATYGLK----- 489  
 Qy 587 YNIGVKTTELNSDGTSDKFSVKGSGTNSISLVTAEHLASLYNEVNRADTSALQSFTVKE-- 644  
 Db 490 -----KVNTAESALQTFVTKKVD 508  
 Qy 645 ---EDDDANAITVAKDTTKNAGAVSLKLGKNGLTVAATKDKGTTFGLSQSDSLTIG 701  
 Db 509 KNGNDANSKIITVGNKKNKPDGTQVNTLKLKGENGVDTTETNGTTFGLNQNGUTVGN 568  
 Qy 702 STLNNDLGLTVKOT--NEQIOVGANGIKFTNVNGSNPOTGIANTRITR----- 747  
 Db 569 STLNNDLGLSVKNTNSKQIOVGADGITFTDLSNPKFAGIENTRITRDRGIGFANNTGSL 628  
 Qy 748 ----- 747  
 Db 629 DANKPRLTPTGINAGGKELTVQSAINPATNGQGLDFMRLSTANTEKSGSAATIKDLYN 698  
 Qy 748 ----- 747  
 Db 689 LSQVPLTFAGDTGNVTKLGEILKVKGGKTTADDLTKNIGVVADSTDNSLTVKLAKTL 748  
 Qy 748 ----- 747  
 Db 749 SLDLAVNTKLTASDKVTVDGSGNNTAKLQNGDLTFSKONTGATPATNSKTIYGVDLKFT 808  
 Qy 748 -----DKIGFAGSGDGVNTKPYLDQDKLOVGNKVTINTGINAGGKAITG 792  
 Db 809 DNNGIALDGTFTYITKDKVGFAGQDGLSKPYLDKDKLVKGEVETITNGINAGGKAITG 868  
 Qy 793 LSPITLPSADQSSRNIELGNTIQDKRKSNAASINDILNTGFNLKNNNPIDFVSTYDID 852  
 Db 869 LSNLTLDATNATTVCHVTQGLIGVSDTDRASIGDVLNAGFNKLNKGDAKDFVSTYDID 928  
 Qy 853 FANGNATTVTHDTANKTSKVYDVVNDVTTIHLTGTDNKK-KLGKVTTLKNTSANGN 911  
 Db 929 FINGNATTVKTYD--GKASKVAYDVNVDDGTTHLTGADGNKNGIKGVKTTTLTKTDKAGD 986

Qy 912 TATNFNVNSSDEDLVNAKDIAENLNTLAKETHHTTKGTADTALQTFVTKVYKVDENNAD-D 970  
 Db 987 KAINFVNSGGDKALINAKDIADNLNTLAGEIRNTKGTADTALQTFVTKVYKENGDDND 1046  
 Qy 971 ANAITVQKNANNQVNTLTKGKGLNTKTOKNGTFTVGTINTTSLKAG-KSTLNDGSL 1029  
 Db 1047 ADTITVGDAKTNOVNTLKLKNGLDIOTNKGDTFTVGTINTTSLKAG-KSTLNDGSL 1106  
 Qy 1030 KNPPTGSOIOVGADGVKFAKNNVNGVVGAGIDGTRITRDRBEIGFTNGSLDKSKPHLS 1089  
 Db 1107 KNTAGNEQIOVGADGVKFAKNNVNGVVGAGIDGTRITRDRBEIGFTNGSLDKSKPHLS 1165  
 Qy 1090 KGINAGGKKTINIQSGETIAQNSHDVATGGKIYDLKTELENKISSTAKTQNSLHEFSVA 1149  
 Db 1166 KGINAGGKKTINIQSGETIAQNSHDVATGGKIYDLKTELENKISSTAKTQNSLHEFSVA 1225  
 Qy 1150 DEQGNFTVSNPYSYDTSKTSVLTFAENGITTKVNGVVRVGDIDTGLTTPKLTVG 1209  
 Db 1226 DEQGNFTVSNPYSYDTSKTSVLTFAENGITTKVNGVVRVGDIDTGLTTPKLTVG 1285  
 Qy 1210 NNGKGIVIDSONGQNTITGLSNTLANVTNDRKSVRTEQGNIIKDEKTRAASIVDVL 1269  
 Db 1286 NNGKGIVIDSONGQNTITGLSNTLANVTNDRKSVRTEQGNIIKDEKTRAASIVDVL 1345  
 Qy 1270 AGFNLOGNEAVDFVSTYDVTNFAAGNATTAKVTYDDTSKTSKVYDVNVDDTTIEVKDK 1329  
 Db 1346 AGFNLOGNEAVDFVSTYDVTNFAAGNATTAKVTYDDTSKTSKVYDVNVDDTTIEVKDK 1405  
 Qy 1330 KLGVTTLTSTGTGANKFALSNQATGDALVKASDIVAHLNTLSGDIOTAKGASQANNSA 1389  
 Db 1406 KLGVTTLTSTGTGANKFALSNQATGDALVKASDIVAHLNTLSGDIOTAKGASQANNSA 1465  
 Qy 1390 GYVDADGNKVIYDSTDNKYIQAANDGTVDKTEKAVDKLVQAQTPDGTGLAQMNVKSVIN 1449  
 Db 1466 GYVDADGNKVIYDSTDNKYIQAANDGTVDKTEKAVDKLVQAQTPDGTGLAQMNVKSVIN 1525  
 Qy 1450 KQVNDANKKOGINEDNAFVKLEKASDNKTNAATVVDGLNVAQTPLTTFAGDTGTGA 1509  
 Db 1526 KQVNDANKKOGINEDNAFVKLEKASDNKTNAATVVDGLNVAQTPLTTFAGDTGTGA 1585  
 Qy 1510 KKLGETLTIKGGTDTNKLTDNNIGVWAGTGTFTVKLAKDLTNLSNVNAGGKIDDKGVS 1569  
 Db 1586 KKLGETLTIKGGTDTNKLTDNNIGVWAGTGTFTVKLAKDLTNLSNVNAGGKIDDKGVS 1645  
 Qy 1570 FVDSGQAKANTPVLNANGDLGCKVISNKGKTKDTDAANVQOLNEVRNLLGLGNAGND 1629  
 Db 1646 FVDSGQAKANTPVLNANGDLGCKVISNKGKTKDTDAANVQOLNEVRNLLGLGNAGND 1705  
 Qy 1630 NADGNOVTADIKKDPNSGSSNRTVIKAGTVLGGKGNNDTEKLTATGGIQGVYDKDGNAN 1689  
 Db 1706 NADGNOVTADIKKDPNSGSSNRTVIKAGTVLGGKGNNDTEKLTATGGIQGVYDKDGNAN 1765  
 Qy 1690 GDLNYSNVWTKQDKGSKKALLATYNAAGQNYLTNNPAPAIIDRINEQIRFFHVNDGNQEP 1749  
 Db 1766 GDLNYSNVWTKQDKGSKKALLATYNAAGQNYLTNNPAPAIIDRINEQIRFFHVNDGNQEP 1825  
 Qy 1750 VVOGRNGIDSSASGKHSVAIGFOAKADGGAAVAIGRQTOAGNOSIAIGDNAOATGQOSTA 1809  
 Db 1826 VVOGRNGIDSSASGKHSVAIGFOAKADGGAAVAIGRQTOAGNOSIAIGDNAOATGQOSTA 1885  
 Qy 1810 IGTGNVAGKHSVAIGFOAKADGGAAVAIGRQTOAGNOSIAIGDNAOATGQOSTA 1869  
 Db 1886 IGTGNVAGKHSVAIGFOAKADGGAAVAIGRQTOAGNOSIAIGDNAOATGQOSTA 1945  
 Qy 1870 GSNATISAGTHAGTQAKKSDGTAGTTTAGATCTVKGAGTAVGAVSVGASCAERIQN 1929  
 Db 1946 GSNATISAGTHAGTQAKKSDGTAGTTTAGATCTVKGAGTAVGAVSVGASCAERIQN 2005  
 Qy 1930 VAAGEVSATSDAVNGSOLYKATQSIANATNELDHRITHQENKANAGISSAMAMASMPQA 1989  
 Db 2006 VAAGEVSATSDAVNGSOLYKATQSIANATNELDHRITHQENKANAGISSAMAMASMPQA 2065









RESULT 4  
US-08-409-995-4  
Sequence 4, Application US/08409995  
Patent No. 5646259  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen I.  
APPLICANT: St. Geme III, Joseph W.  
TITLE OF INVENTION: Haemophilus Adhesion Proteins  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/409,995  
FILING DATE: 24-MAR-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-61053/REF  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1912 amino acids  
TYPE: amino acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
US-08-409-995-4

Query Match 7.7%; Score 792.5; DB 1; Length 1912;  
Best Local Similarity 22.7%; Pred. No. 2.3e-40;  
Matches 474; Conservative 239; Mismatches 652; Indels 727; Gaps 99;

Qy 1 MNHYKVFNKATGTFMVAEYAKSHSTGGSCATGQVGS-----VCTLSFARAAALAV- 54  
Db 1 MNKIFNVNVMQTWVWVSELTRTH-----TKRLNRNGDPVLATLLFATVQANATD 52  
Qy 55 -----LVIGATLSSGSAVAKKDKHIAIGRONQPRRS-----GTAKADGDRAIAI 99  
Db 53 EDEELDPVVRAPVLVSFHSDEGT-----GEKEVTENSNGIYFDNKGVLKA-GAIIILKA 106  
Qy 100 GBNAAQGOAIAIGSSNKTNGS-----SLDKIGTADT--QGESIATGGDVKASGD-ASI 152  
Db 107 GDNLKKX-----QXTDEXTNASSFTYSLKKDLTDLTSVATEKLSEF-----ANGDKVDI 155  
Qy 153 AIGSDDLHLLDQHGPKPKG--TLINDLINGHVALEKRSKDNQKYRRT----- 202  
Db 156 TSDANGLKLAKTNGNNGVHLNGLDSTLPDAVNTNGVLSS--SSEFTNDVKEKTRAIVKDVLN 214  
Qy 203 -----TASGHASTAVGMSYAQQHFSNAGFTRAITAKSAYSIAVGLAAATAEQSTIA 253  
Db 215 AGWNINKAGTAGGNVESDLVSAYNNVEFIT--GDKNT-----LDVVLTAENKXKITEV 266  
Qy 254 IGSDATSSSLGALALCAGTRAQGLSIALGQGSVVYTSQSDNNRRPAYTPNTQALDPKFOAT 313  
Db 267 KFTPKTS-----VIKEKDGKLEFTGKENDNTNKVTSNT-----AT 300  
Qy 314 NNTKAGPLSIGNSIKRKLIINVGAGVKNKTDVNVVAQLEAVVGVKAKERRITTFQGDNDNST-- 371

Db 301 DNTDEGNGLVTAKAVIDAVNKAQWRVKTITTTANGONGDFATV--ASGNTVTFESGDGTAS 358  
Qy 372 -----DVKIGLDNT-LTIKGG----- 386  
Db 359 VTKDTNGNGITVKYDAKVGDOGDKFDSOKKIYADVTALTITVTGGKVAETAKEDDKKKLVNAG 418  
Qy 387 -----AETNALTDNNI-GVVKREAD-----NSGLKVK-----LA 413  
Db 419 DLVTFALGNLSWKAKEADTDGALGISKDQEVKAGETVTFKAGKNLKVQDCANFTVSLQ 478  
Qy 414 KTLNMLEVTNTTTLNATTTVKVSSSSTTAEILLSLFTPTOPNTGSSQSTSKTV-YGVNGV 472  
Db 479 DALTLGLTSI--TLGGTTN--GGNDAKTV-INKDGTLTTPAGNGGTGTGTITSVTKDGI 531  
Qy 473 KFTNNAETTAIG-----TTRITRDKIGFARDG--DVDEKQAPLYLDRKQL 515  
Db 532 KAGNKAITNVASGLRAYDDANFDVLNNSATDLNR-HVEDAYKGLLLNLNEKNA--NKQPL 587  
Qy 516 KVGSAITIDN-----GIDAGNKKISNLAKGSSANDAVTIEQLKAAPKPTLNAGAGIS 567  
Db 588 VTDSTAATVGDRLKLGWVSTKNGTKESN-----QVQDADEVLTFCAGAA 633  
Qy 568 VTFPEISYDAKSGNVTAPTYNIGVKTE-----LNSDGTSDKFSVKGSGTNNSLVTAEHLA 623  
Db 634 -----TVTSKSEN-GKHTITVSVAETKADCGLEKDGDTIKLKVDNQNTDNLVTGN--- 683  
Qy 624 SYLNEVNRADTSALQSFTVKE-EDDDANALITVAKDTTKNAGAVSILKLGKNGLTUVATK 682  
Db 684 -----NGTAVTKGGFEIVKTGATDADRGKTV-KDATANDADKKVATVK-----DVA 731  
Qy 683 KQGTVTF-----GLSDSGLTIGKSLNNDGLTVKDT-----NEQIOVGANGIKPT- 728  
Db 732 INSAATFVKTENLTTSIDEDNPTDMCK-----DDALKAGDTLTFKAGKNLKVARDCKNITF 787  
Qy 729 -----NVNGSNPGTGIANTARI-----TRDKIGFAGSDGAYDTNKPYLD 767  
Db 788 DLAKNLEVKTAKVSDTLTIGGNTPTGGTTATPKVNTITSTADGLNFA----- 833  
Qy 768 QDKLOVGNVKIITNTGINAGKAI--TGLSPTL--PSIADQSSRNIELGNTIODKDKSNAA 823  
Db 834 -----KETADASGKNVYLKGIATTLTLEPSAGAKSS-HVDL--NVDATKKSNA 879  
Qy 824 SINDILNTGENLKNNNPIDFVSTYDIDVFANGNATATVTHDTANKTSKVVDVNVDDT 883  
Db 880 SIEDVLRAGNIQNGNNVDYVATYDVTNFTDDSTGTTTIVT-----V 921  
Qy 884 TIHLTGTDDNKKLGKVTTKLNKTSANGNTAT-----NFNVNSDED-----ALVNAK 930  
Db 922 TOKADGKADVKIGAKTSVIK--DHNGKLTGKDLKDANNGATVSEDDGKDTGTGLVTAK 979  
Qy 931 DIAENLNTLAKEIHTTKGTADTALQTFYVKVDENNADANAITVGO---KNANNQVN 986  
Db 980 TVIDAVNKGSRVTFEGATAETGATAV-----NAGNAETVTSVNFKNGN--AT 1028  
Qy 987 TLTLLKENG--LNITDKNGTIVTFEINTTSLKAG-----KSTLNDGGLSTKNPTGSE 1037  
Db 1029 TATVSKDNGNINKYD-----VNVGDLKIGDDKKIVADTTTLTVTGGKVSVPAGAN 1080  
Qy 1038 QIQ-----VGADGVKFAKVN----- 1053  
Db 1081 SVANNKKLVNAEGLATA-LNNLSWTAKADKADYADGESEGTQDEYKAGDKVTFKAGKNLV 1139  
Qy 1054 -----NCVVGAGIDGT-----TRTRDEIGTGTNGS---LDKSKPH 1087  
Db 1140 KQSEKDPYTSLODTLTGLTSITLGGTANGRNDGTIVLNKDLGITLANGAAAGTASNGN 1199  
Qy 1088 ---LSKDGINAGKKTINIOG---EIAONSHDADVTKGIYDLKTELENKI----- 1132  
Db 1200 TISVTKDGISAGNKEITNVKSALKTYKDTONTADE-TQDREFHAAVKNAEVEFVGKNGA 1258  
Qy 1133 SSTAKTAQNSLIH-----EFSVADEQGNFTVSNPYSSVDTSK 1169



Db 684 -----NCTAVTKGFFETVGTATDGRKVTV-KDATANDAKKATVK-----DVATA 731  
QY 683 KOGTVTF-----GLSQDSGLTIGKSTLNNDGLTVKDT-----NBOIQVGANGIKFT- 728  
Db 732 INSAATPVKTENLTTSIDENPDNCK-----DALKAGDVLTFKAGKLNKVKRDKNITF 787  
QY 729 -----NVGNSRPGCIANTARI-----TRDKTGFAGSDCAVDTKRKYLD 767  
Db 788 DLAKNLEVKAKVSDTLTICGNTPTGGTTATPKVNIITSTADGLNFA----- 833  
QY 768 QDKLQGVNWKITNGINAGKAL--TGLSPTL--PSIADOSSRIELNFIQDKKSNA 823  
Db 834 -----KETADASGSKNVLKGIATLTLEPSAGAKSS-HVDL--NVDATKKSNA 879  
QY 824 SINDILNTGNLNNNNPIDFVSTYDIDVFANGNATATVTHDTANKTSKVVDVNVDDT 883  
Db 880 SIEDVLKRNIOGNGNVYVATYDVTNFTDDSTGTTVT-----V 921  
QY 884 TIHLTGDDNKKLGCVTKTKLNKTSANGNTAT-----NPNVSSDED-----ALVNAK 930  
Db 922 TQADGKGVADKIGAKTSVTK--DHNGKLTGDKLDANNAGATVSEDDGKDTGTGLVTAK 979  
QY 931 DIAENLTLAKEIHTTKGTADTALQTFVKKVDENNADANAITVQO-----KNANQVN 986  
Db 980 TVIDAVNKSCHWRTGEGATETATAV-----NAGNAETVTSCTSVNFRGN--AT 1028  
QY 987 TLLKKGNG-LNITKDKNGVTVEINTTSLKAG-----KSTLNDGGLSKNPTGSE 1037  
Db 1029 TATVSKDNGNINKYD-----VNVGDLKIGDDKKIVADTTTLTGTGKVSVPAGAN 1080  
QY 1038 QIQ-----VCADGVKFAKVN----- 1053  
Db 1081 SVNNKKLVNAEGLATA-LNLSWTAKADYADGESEGTQDEVKAGDKVTFKAGNKKV 1139  
QY 1054 -----NCVVGAGIDGT-----TRTRDEIGFTGNGS-----LDKSKPH 1087  
Db 1140 KQSEKDFYSLQDTLGLTITIGTGANGRNDTGTWINKDCLTITLANGAAAGTDSNGN 1199  
QY 1088 -----LSKDGINAGKKTNIQSG-----EIAQNSHDVATGGKIYDLKTELENKI----- 1132  
Db 1200 TISVTKDGISAGKEITNVKSAKTYKDTQNTADE-TQDKFHAHVKNANEVEFVGKNGA 1258  
QY 1133 SSTAKTAQNSLH-----EFSVADEQGNFTVSNPNSSYDTSK 1169  
Db 1259 TVSAKTDNNGKHVTIIVDAEAKVGDGLEKDTGDKIKLKVDNTDGNLLTVDATKGASVAK 1318  
QY 1170 -----TSDVITPAGENG-----ITTKVNGVVRVGIDOTFKGLT----- 1202  
Db 1319 GEFNAVTTDATTAQGINANERGVVKGSGNGATATETDKKKVATGVDVAKAINDAATFVK 1378  
QY 1203 ----- 1202  
Db 1379 VENDDSATIDSDPTDGDANDALKAXDTLTLKAGKLNKVKRDKGNITFALANDLSVKSATV 1438  
QY 1203 TPKLTVCNNNGK-----GIVDTSQNGNT-----ITCLSNLTANVNDKGSVRTTEQ 1249  
Db 1439 SDKLSLGTNGKNVNIISDTKGLNFADKSKGTGDDANTHLNGIASTLTDTLLNSGAT-TNLG 1497  
QY 1250 GNLIKEDKTRAASIVDVLSAGENLQO-----NGEAVDFVSTYDVTNFDAGNATTA 1301  
Db 1498 NGGITDNEKXAASVKDVLNAGVNRGVRPASANNQVENIDFVATYDVTDFVSGDKDPTS 1557  
QY 1302 VTYD--DTSKTSKVVDVNVDDTTIEVKOKKLGKVTTLTSTG-----TGAN-KFALSNOA 1354  
Db 1558 VTVESKDNGKRTVE-----KIGAKTSVKDHNGKLTGKELKDANNNGV 1601  
QY 1355 T-----GDALVKASDILVAHLNLTSGDIOTAKASQANNSAGYVDAGKNKIYDSTD 1405  
Db 1602 TVTETGKDEGNGELVTAKAVIDAVNAGVRVKT--GANGQNDFAFV--ASGTNVT-- 1655  
QY 1406 NKYYOARNDGTVDKTEKAVARDKLVAQAQTPDGTPL-AQMNVKSVINKEQVNDANKKOG--- 1461

Db 1656 -----ADGNGT---TAEVTK-----ANDGSITVKYNVK-----VADGLKLDGDKI 1692  
QY 1462 INEDNAFVKGLEKAASDN-----KTKNAAVTVGDLNNAVAQTPLTFAGDGTGT----- 1507  
Db 1693 VADTTLTVADGVKVTAPNNGXGKFKFXDASGLAGCLNKLST--ATAGKEGTGEVDPANSA 1750  
QY 1508 -TAKKLGELTITIKGGTDDTNKLDNNGIIGVVAGTGDGFTVVKLAKDLTLNLSV-----NAG-- 1559  
Db 1751 GQEVKAGDKVTFRAG-----DNLKIKQSKDFTYSLKKELKDLTSVEFKDANGGTG 1801  
QY 1560 --GTKIKDDKGVSVFSSGQA-----KANTPVLSSANGLDLGGKVISNVGKGTK 1604  
Db 1802 SESTKITKDLGLTTPANGACAAGANTANTISVTKDGLISAGNKAVTNVVSGLK 1853  
RESULT 6  
US-08-409-995-2  
; Sequence 2, Application US/08409995  
; Patent No. 5646259  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen I.  
; APPLICANT: St. Gene III, Joseph W.  
; TITLE OF INVENTION: Haemophilus Adhesion Proteins  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/409,995  
; FILING DATE: 24-MAR-1995  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-61053/RFT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1098 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; US-08-409-995-2

Query Match 6.6%; Score 678.5; DB 1; Length 1098;  
Best Local Similarity 23.6%; Pred. No. 1.1e-33;  
Matches 304; Conservative 157; Mismatches 428; Indels 399; Gaps 50;  
QY 944 HTTKGTADTA---LQFTTVKKVDENNADANAII-TVGKNANNQVNTL-----TL 990  
Db 26 HTKASATVAVATLALLSATVEANNNTPTVTKLKAYGDANFNFTNNSIADEAKQVOEAY 85  
QY 991 KGBNGLNINIKTDKNGTTFGTGINTTSGLKAGKSTLNDGGLSKNPTGSEIQVGADGVKFAK 1050  
Db 86 KGLLNLN---EKNASDKLLVEDNTAATVGNLRLKGLWLSKNGKTRNEKSO----QVKAH- 137  
QY 1051 VNNNGVVGAGIDGTTTRITRDEIGFTGTGNSLDKSKPHLSKDGINAGGKKTITNTQSGEIAQ 1110  
Db 138 -----DEVLFEG-----KGVQVTS----- 152

QY 1111 NSHDAVTGGKIYDLKTELENKISSITAKTAQNSLH--EFSVADEOQNFTVSNPSSYDTS 1168  
Db 153 -----TSNGKHITTFALAKDLG-----VKTA 174  
QY 1169 KTSVITFAG--ENGITTKVNGVVRVIGDQTKGLTTPKLTIV--GNNGKGIVIDS--QNG 1223  
Db 175 TVSDTLFVGGAACAT-----TTPKVNVTSTTDLGKFAKDAAGANG 216  
QY 1224 QNTI--TGLSNTLANVTNDKGSVRTTEQGNLIKDEDK--TRAASIVDVLSAGFNQ----- 1275  
Db 217 DTTVHLNIGISTLDTL--VGSPTHIDGG--DOSTHYTRAASIKDVLNAGWNKGVKA 271  
QY 1276 ---GNGEAVDFVSTYDTVNFADGNATATKVTYDDTSKTSKVYVDVNVDDTTIEVKDKKL 1331  
Db 272 GSTTGQSENVDVHTYDTVEFLSADTETTTVDSKENGKRTEYKIGAKTSVKEKDGKL 331  
QY 1332 GV-KTTTLTSTGTGANKFALSNOATGALVKASDIVAHLNLTSGDIDOTAKGASOANSAG 1390  
Db 332 FTGKANKETHKVDGAN--ATEDADEGKGLVTAKDVIDAIVNKTGWRIKVTTDANGONGDFAT 389  
QY 1391 YVD-----ADGN--KVIVYDSTDKYQAKNDGTVDKTEVAKDKLVAQAQTPDGLTIAQ 1441  
Db 390 VASGTNVTFASGNCTTATVINGTDG--ITVKYDAKVGDLKLDGDKIAA-----DTTALT 442  
QY 1442 MNKSVINKEOVNDANKKOGINEDNAPVKGLEKAASDNK---TKNAAVTVGDLNVAQTP 1498  
Db 443 VN-----DGKNANPRKGVADVASTDEKKLVTAKGLVTA--LNSLSWTT 484  
QY 1499 LTFAGDTG-----TAKKLGELTITKGGOTDNTKLTDDNNIGVAGTGDGTVKLAKDLT 1551  
Db 485 TAAADGGTLTLDGNASEQEVKAGDKVTFKAG-----NLKVKQGANFTYSLQDALT 535  
QY 1552 NLNSV-----NAGTKIDKGVFSVDSGQA--KANTPVLANSGLDGGKVISNVGKGT 1603  
Db 536 GLTSITLGTGNGAKTEINKDGLTITPANGAGANNANTISVTKDGSAGSGSVKVVYSGL 595  
QY 1604 KDTDAANVQOL-----NEVRNLGLGNAGDN-----ADGNQVNIAD----- 1640  
Db 596 KKFGDANFDPLTSSADNLTKONDDAYKGLTNLDEKGTDKQTPPVVADNTAATVGLRLGLG 655  
QY 1641 -IKKDPNSGSS-----NRTVTKAGTVLGGKGNNDTEKLTATGGIOVGVDPKDGANGDL 1692  
Db 656 VISADKTTGGSTFHDQVRNANEVK---FKSGNGINSGKTVNGRREITFEL---AKGEV 709  
QY 1693 --SNWVKTKDQSGKALLAT---YNAAGQTNLYTNPAEAIIDRINEQGRFFHFVNDGNQ 1747  
Db 710 VKSNEFTVKETNGKETS LVKGVKYSKEDIDLTTGP-----NLKDG- 753  
QY 1748 EPVVOGRNGIDSSAGSKHSVAIGFOAKADGEAAVAIGRQTOAGNQSIAIGDNAQATGDQS 1807  
Db 754 -----TVAAKYQDK-----GGKVVSVTDNTEATITNK 780  
QY 1808 IAGTGNVAGKHSGAIGDPSTVKRADNSYSVGNNO-----FTDATQTDVFGVGNNTIV 1861  
Db 781 ---GSGVTGQVADAL-----AKSGFELGLADEADAKRAFODKTKALSAGTTEIYNA 830  
QY 1862 TESNSVALGNSAISAGTHAGTQAK-----KATQSIANAT 1959  
Db 831 HDKVFREANGLNTKYSAAVTESTDANGDKVTTTFVKTVDVELPLTOIYNTDANGKKITKVKV 890  
QY 1887 -----KSDGTAGTTTT-----AGATGVTKGFAGQTAAG 1914  
Db 891 DGQTKWYELNADGTADMTKEYTGLNVSDGKGVKVDNDGKWHAKADGTADTKRGEVSD 950  
QY 1915 AVSVG-----ASGAERLQNVAAGEVSAVSTSDAVNGSOLY---KATQSIANAT 1959  
Db 951 KVSDEKHVWSLDPNDQSGKGVVDNVANGDISATSDAINGSOLYAVAGVYTNLAGQV 1010  
QY 1960 NELDRHONENKANAGISSAMAMASPMQAYIPGRSMVVTGGIATHNGOGAVAGLSKLS 2019  
Db 1011 NNLSEKVNKVKRADAGTASALASOLPQATMPKGSWVAIAGSSYQONGLAIGVSRISD 1070  
QY 2020 NGQWFKINGSADTQGHVGAAGACPHF 2047

Db 1071 NGKVIIRLSGTTNSQKGTGVAAGVGYQW 1098

## RESULT 7

US-08-685-467-2  
: Sequence 2, Application US/08685467  
: Patent No. 6060059  
: GENERAL INFORMATION:  
: APPLICANT: St. Gene III, Joseph W.  
: APPLICANT: Barenkamp, Stephen J.  
: TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
: NUMBER OF SEQUENCES: 6  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
: STREET: Four Embarcadero Center, Suite 3400  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: United States  
: ZIP: 94111-4187  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/685,467  
: FILING DATE: 22-JUL-1996  
: CLASSIFICATION: 424  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/409,995  
: FILING DATE: 24-MAR-1995  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Silva, Robin M.  
: REGISTRATION NUMBER: 38,304  
: REFERENCE/DOCKET NUMBER: A-61053-2/RPT/RMS  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (415) 781-1989  
: TELEFAX: (415) 398-3249  
: TELEX: 910 277299  
: INFORMATION FOR SEQ ID NO: 2:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1098 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: unknown  
: TOPOLOGY: unknown  
: MOLECULE TYPE: DNA (genomic)  
US-08-685-467-2

Query Match 6.6%; Score 678.5; DB 3; Length 1098;  
Best Local Similarity 23.6%; Pred. No. 1.1e-33;  
Matches 304; Conservative 157; Mismatches 428; Indels 399; Gaps 50;

QY 944 HTTKGTADTA---LQTFVKKVDENNADANAI-TVGOKNANNQVNTL-----TL 990  
Db 26 HTKASATVAVAVLATLLSATVEANNPTVTKLAKYDANFNFTNNSIADAQKQVEAY 85  
QY 991 KGENCLNKTDKNGTVTEGINTTSLGKAGKSTLNDGGLSIKNPTEGSEQIOVGADGVKFAK 1050  
Db 86 KGLLNLN---EKNASDKLLVEDNTAATVGNLRKLGWLVSSKNGTRNEKSQ-----QVKA- 137  
QY 1051 VNNGVVAGIDGTTTRITRDEIGFTGTNGSLDKSPHLSKDGINAGKKITNIQSGEIAQ 1110  
Db 138 -----DEVLFEG-----KGQVQVTS----- 152  
QY 1111 NSHDAVTGGKIYDLKTELENKISSITAKTAQNSLH--EFSVADEOQNFTVSNPSSYDTS 1168  
Db 153 -----TSNGKHITTFALAKDLG-----VKTA 174  
QY 1169 KTSVITFAG--ENGITTKVNGVVRVIGDQTKGLTTPKLTIV--GNNGKGIVIDS--QNG 1223  
Db 175 TVSDTLFVGGAACAT-----TTPKVNVTSTTDLGKFAKDAAGANG 216

```
QY 1224 QNTI--TGLSNTLANVNDKGSVRTTEQGNIIKDEK--TRAASIVDVLSAGFNLO----1275
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 217 DTTVHLNGIGSTLTDTL--VGSPTHIDGG---DQSTHVTTRAASIKDVLNAGWNIKGVA 271
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1276 ----GNGEAVDFVSTVDYTNFADGNATTAKYVYDDTSKSVYDVYVNDVDTTIEVKDKKL 1331
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 272 GSTTGOSENVDFVHTYDTVEFLSADTETTTVTVDSEKNGKRTVEKIGAKTSVKEKDGKL 331
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1332 GV-KTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTLSGDIOTAKGASOANNSAG 1390
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 332 FTGKANKETNKVDGAN--ATEDADECKGLVTAKDVIDAVNKTGWRKTTDANCQNGDFAT 389
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1391 YVD-----ADGN---KVIYDSTDNKYQAKNDGTVDKTEKAVAKDLVAAQAQTPDGTLLAQ 1441
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 390 VASGTVNVTFAASGNTTATVNTGTDG--ITVKYDAKVGDLGLKLDGDKIAA-----DTTALT 442
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1442 MNVKSVINKEQVNDANKKOGINEDNAFVKGLEKAASDNK---TKNAAVTVDGLNVAQTP 1498
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 443 VN-----DGKNANPNKGVADVASTDEKRLVTAKGLVTA--LNSLSWT 484
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1499 LTFAGDTGT-----TAKKLGTELTIKGGQTDNTKLTNNIGVAGTDTGFTVKLAKDLT 1551
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 485 TAAEADGGTLDGNASEQEVKAGDKVTFKAG-----NLKVKQEGANFTVSLQDALT 535
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1552 NLSNV-----NAGGPKIDDKGVSVFVDSGQA--KANTPVLSSANGLDLGGKVISNVGKGT 1603
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 536 GLTSITLGTGNNGAKTEINKDGLTITPANGAGANNANTISVTKDGISAGSQSVKVVYSGL 595
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1604 KDTDAANVQOL-----NEVRNLLGLGNAGNDN---ADGNQVNTAD-----1640
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 596 KKEPDANFDFLSSADNLTKQNDAYKGLTNLDEKTDKQTPVVAWNTAATVGDGLRGLGW 655
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1641 -IKKPNSSGSS-----NRTVIKAGTVLGGKGNNDTEKLTATGGIQGVGDKNAGNDL 1692
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 656 VISADKTTGGSTVEYHQPVRNANEVK---FKSGNGINVSQKTVNGRREITFEL---AKGEV 709
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1693 --SNVWVKTKDGSKKALLAT---YNAAGQTNVLTNNPAAEDRINEQIRFHVNDGNQ 1747
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 710 VKSNEFTVKTNGKETSLSKVGKDYKYSKEDIDLTTCQP-----KLKDGK- 753
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1748 EPVVGQRNGIDSSASGSHVAIGFOAKADGEAAVAIGROTQAGNQSIATGDNAQAATGDQS 1807
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 754 -----IYAAKYQDK-----GGKVVSVTDNTEATITNK 780
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1808 IATGTGNVAGKHSAGIDPSTVKADNSYSVGNNO-----FTDATQTDVFGVGNNTIV 1861
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 781 ---GSGVYTGNOVADAI-----AKSGFELGLADEADAKRAPDDTKALSAGTTEIVNA 830
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1862 TESNSVALGNSAISAGTHAGTOAK-----1886
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 831 HDKVRFANGLNTKVSATVESTDANGDKVTTTFVKYTDVDELPLTQIYNTDANGKKTIVKVK 890
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1887 -----KSDGTAGTTTT-----AGATGTIVKGFAGOTAVG 1914
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 891 DGOTKWVELNADGTADMTKEVTLGNVDSGKKVYKNDGDKWYHAKADGTADTKTGEVND 950
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1915 AVSVG-----ASGAERIRIONVAAGEVSATSTDAVNGSOLY----KATQSIANAT 1959
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 951 KVSDEKHVVSLDPNPQSKGKGVVDNANGDISATSTDAINGSOLYAVAKGVNTLAGQV 1010
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1960 NELDRIHONENKANAGISSAMAMSPQAYIPGRSMVTGGIATHNGOGAVAVGLSKLSD 2019
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 1011 NNLGKVNKVGKRADAGTASALAAASOLPOATMPCKSMVAIAGSSYOGQGLAIGVSRISD 1070
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 2020 NGQWPKINGSADTOGHVGAAGVAGFHF 2047
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 1071 NGKVIIRLSGTTNSQKGTGVAAGVGYOW 1098
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
```

RESULT 8

US-09-377-155-32

; Sequence 32, Application US/093771155

```
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 085064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 1098
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-377-155-32
```

```
Query Match 6.6%; Score 678.5; DB 4; Length 1098;
Best Local Similarity 23.6%; Pred No 1.1e-33;
Matches 304; Conservative 157; Mismatches 428; Indels 399; Gaps 50;
```

```
QY 944 HTTKGTADTA---LQFTTVKVDENNNDADANAI-TVGOKNANNQVNTL-----TL 990
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 26 HTKASATVAVAVLATLLSATVEANNPTVTKLKYAGDANFTNNSADAEKQVQEAY 85
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 991 KGENGLNIKTNGTTFGINTTSGLKAGKSTLNDGSLTKKNTPTGSEQIQVGADGVKFAK 1050
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 86 KGLLNIN---EKNASDKLLVEDNTAATVGNLKRKLGVLSKNGTRNEK90---QVKHA- 137
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1051 VNNNGVVGAGIDGTTTRITREIGFTGTNGSLDKSPHLSKDGINAGGKKTITNQSIEIAQ 1110
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 138 -----DEVLFEG-----KGVQVTS-----152
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1111 NSHDAVTVGKIYDLKTELENKISSTAKTAONSUHL--EFSVADBOGNNTVSNPYSYDTS 1168
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 153 -----TSENGKHTTTFALAKDLG-----VKTA 174
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1169 KTSDDVITFAG--ENGITTKVNGVVRVGIDQTKGLTTPKLTV-GNNGKGIVIDS--QNG 1223
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 175 TVSDTLITICGGAAGAT-----TPPKVNVTSITDGLKFAKDAAGANG 216
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1224 QNTI--TGLSNTLANVNDKGSVRTTEQGNIIKDEK--TRAASIVDVLSAGFNLO----1275
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 217 DTTVHLNGIGSTLTDTL--VGSPTHIDGG---DQSTHVTTRAASIKDVLNAGWNIKGVA 271
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1276 ----GNGEAVDFVSTVDYTNFADGNATTAKYVYDDTSKSVYDVYVNDVDTTIEVKDKKL 1331
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 272 GSTTGOSENVDFVHTYDTVEFLSADTETTTVTVDSEKNGKRTVEKIGAKTSVKEKDGKL 331
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1332 GV-KTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTLSGDIOTAKGASOANNSAG 1390
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 332 FTGKANKETNKVDGAN--ATEDADECKGLVTAKDVIDAVNKTGWRKTTDANCQNGDFAT 389
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1391 YVD-----ADGN---KVIYDSTDNKYQAKNDGTVDKTEKAVAKDLVAAQAQTPDGTLLAQ 1441
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 390 VASGTVNVTFAASGNTTATVNTGTDG--ITVKYDAKVGDLGLKLDGDKIAA-----DTTALT 442
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1442 MNVKSVINKEQVNDANKKOGINEDNAFVKGLEKAASDNK---TKNAAVTVDGLNVAQTP 1498
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 443 VN-----DGKNANPNKGVADVASTDEKRLVTAKGLVTA--LNSLSWT 484
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1499 LTFAGDTGT-----TAKKLGTELTIKGGQTDNTKLTNNIGVAGTDTGFTVKLAKDLT 1551
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 485 TAAEADGGTLDGNASEQEVKAGDKVTFKAG-----NLKVKQEGANFTVSLQDALT 535
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1552 NLSNV-----NAGGPKIDDKGVSVFVDSGQA--KANTPVLSSANGLDLGGKVISNVGKGT 1603
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 536 GLTSITLGTGNNGAKTEINKDGLTITPANGAGANNANTISVTKDGISAGSQSVKVVYSGL 595
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
```

QY 1604 KDTDAANVOQL-----NEVRNLLGLGNAGNDN-----ADGNQVNIAD----- 1640  
Db 596 KKFGDANFDPPLTSSADNLTKONDDAYKGLTNLDEKGTDKOTPPVADNTAATVGLRGLGW 655  
QY 1641 -IKDPNSGSS-----NRTVTKAGTVLGGKGNNDTEKLGATGGIOGVVDKDCGNANGDL 1692  
Db 656 VISADKTTGGSTVHDQVRNANEV-----FKSGNGINVGKTVNGRREITFEL---AKGEV 709  
QY 1693 --SNVYKTKQDKSKALLAT---YNAAGTNYLTNNPAEAIDRINEQGRFFHFVNDGNQ 1747  
Db 710 VKSNEFTVKETNGKETS LVKVDKRYYSKEDIDLTTGP-----NLKDCN- 753  
QY 1748 EPVVGNGRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTAGNQSTAIGDAAOAGDQS 1807  
Db 754 -----TVAARYQDK-----GGKVVSVTDNTEATITNK 780  
QY 1808 IAITGNVAGKHSAGTDPSTVKADNSVSGNNNQ-----FTDATQDVFVGNNITV 1861  
Db 781 --GSGVYTGNOVADAI-----AKSGFELGLADEADAKRAFDDKTKALSAGTEIVNA 830  
QY 1862 TESNSVALGNSAISAGTHAGTOAK-----AGATGTVKGFAGOTFAVG 1914  
Db 831 HDKVRFANGLNTKVSAAATVESTDANGDKVTTTFVKTDVELPLTOIYNTDANGKKITKVK 890  
QY 1887 -----KSDGTHAGTTTT-----ACATGTVKGFAGOTFAVG 1914  
Db 891 DGQTKWYELNADGTADMTKEVTGLNVDSDGKKVVKDNDGKWYHAKADGTADKTRGEVSN 950  
QY 1915 AVSYG-----ASGAERRIQNVAAEVSATSTDAVNGSOLX---KATOSIANAT 1959  
Db 951 KYSTDEKHVSLDPNDOSKGGVVINDVANGDISATSDAINGSOLXAVAKGVTNLAGOV 1010  
QY 1960 NELDRIHQENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGQGAVALGSLKSLD 2019  
Db 1011 NNLGKVKNGVKRADGATASALAASQLPQATMPCKSMVAITAGSSYQONGLAIGVSRISD 1070  
QY 2020 NGQWFKINSADTQGHVGAAGVAGFHF 2047  
Db 1071 NGKVIIRLSGTTNSQGRKTGVGAAGVGYQW 1098

## RESULT 9

US-08-913-942-2  
; Sequence 2, Application US/08913942  
; Patent No. 6200578  
; GENERAL INFORMATION:  
; APPLICANT: St. Gene, Joseph  
; APPLICANT: Batenkamp, Stephen J.  
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,942  
; FILING DATE: 29-DEC-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/409,995  
; FILING DATE: 24-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/4031  
; FILING DATE: 22-MAR-1996

; ATTORNEY/AGENT INFORMATION:  
; NAME: Vance, Dolly A.  
; REGISTRATION NUMBER: 39,054  
; REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1098 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-08-913-942-2

Query Match 6.6%; Score 678.5; DB 4; Length 1098;  
Best Local Similarity 23.6%; Pred. No. 1.1e-33;  
Matches 304; Conservative 157; Mismatches 428; Indels 399; Gaps 50;  
QY 944 HTTKGTADTA---LQTFVKKVDENNADANAI-TVGOKNANNQVNTL-----TL 990  
Db 26 HKCASATVAVAVLATLLSATVEANNPTVTKKAYGDANFNNTNSDAEAKOVQBEAY 85  
QY 991 KGENGLNKTDKNGFTVTFTGINTTSGLKAGKSTLNDGGLSINKNPTGSEQIQVGADGVKFAK 1050  
Db 86 KGLLENL---EKNASDKLLVEDNTAATVGNLRKGLWVLSKNGTRNEKSQ---QVKAH- 137  
QY 1051 VNNGVVAGIDGTTTRTRDEIGFTGTNGSLDKSKPHLSKDGINAGGKKIINQSGEIAQ 1110  
Db 138 -----DEVLEFG-----KGVQVTS----- 152  
QY 1111 NSHDAVTGGKIYDLKTELENKISSAKTAQNSLH---EFSVADEQGNNTFTVSNPYSSYDTS 1168  
Db 153 -----TSENGKHITITFALADLG-----VKTA 174  
QY 1169 KTSVDITFAG---ENGITTKVNGVVRVVGIDQTKGLTTPKLTV-GNNGKGVIDS---ONG 1223  
Db 175 TVSDTLTIGGGAAGAT-----TTPKVNVSTTDGLKFAKDAAGANG 216  
QY 1224 QNTI---TGLSNTLANVNDKGSVTRTEOGNIKKDEK---TRAASIVDVLSAGFNLO--- 1275  
Db 217 DTTVHLNGIGSTLDTL--VGSPTHIDGG---DQSTHYTRAASIKDVLNAGWNKGVKA 271  
QY 1276 ---GNGEAVDFVSTYDVFNFADGNATTAKVYDDTSKTSKVVDVNVYDDTTIEVKDKKL 1331  
Db 272 GSTTGQSENVDVFVHTYDVFELSADTETTTVTVDSEKNGKRTVEKIGARTSVIREKDGKL 331  
QY 1332 GV-KTTTLTSTGTGANKPALSNQATGDALVKASDIIVAHNLTLGSDIQTAKGASQANNSAG 1390  
Db 332 FTGRANKETNKVDGAN--ATEDADEGKGLVYAKOVIDAVNKTGWRIKTTDANGQNGDAT 389  
QY 1391 YVD-----ADGN---KVIYDSTONKYYOAKNDGVIVDKTEKVAKDLVAAQATPDGTIAQ 1441  
Db 390 VASGTNVTFAASGNTTATVNTGTDG--ITVKYDAKVGDLKLDGDKIAA-----DTALT 442  
QY 1442 MNKSVINKEQVNDANKKQINEDNAFVKGLEKAASDNK---TKNAAVTVDNLNAVATP 1498  
Db 443 VN-----DGKNANNPKGVADVASTDEKKLVTAKGLVTA---LNSLSWTT 484  
QY 1499 LTFAGDTGT-----TAKKLGETILT IKGGQTDTKNLTDNNIGVVGAGTDFGVTKLAKDIT 1551  
Db 485 TAAEADGGTLOGNASEQEVKAGDKVTFKAGK-----NLKVKQEGANFTYSLODALT 535  
QY 1552 NLNSV-----NAGGTKIDDKGVSVFVSDSSGQA--KANTPVLISANGLDGGKVISHWGKT 1603  
Db 536 GLTSITLTGTGNGNKAITEINKDGLTTPANGAGANNANTISVTKDGISAGGQSVKNVWSGL 595  
QY 1604 KDTDAANVOQL-----NEVRNLLGLGNAGNDN-----ADGNQVNIAD----- 1640  
Db 596 KKFGDANFDPPLTSSADNLTKONDDAYKGLTNLDEKGTDKOTPPVADNTAATVGLRGLGW 655





Qy 1902 GTVKGAGQAVGAVSGASGERRIONVAAGEYSATSTDAVNGSQLY---KATQSTANA 1958  
Db 538 -----DADKIKVASDGISAGNAKAVNVAAGEISATSTDAINGSQLYAVAKVNTNLAQ 590  
Qy 1959 TNELDHRTHONENKANAGISSAMAMSPQAYIPGRSMVTGGIATHNGOGAVAVGLSKLS 2018  
Db 591 VNNLEGVKNVKKRADAGTASALAASQLPQATMPGKSWSTAGSSYQGGNGLAIGVSRIS 650  
Qy 2019 DNGQVFKINGSADTQGHVGAAGAGHFH 2047  
Db 651 DNGKVIIRLSGTTNSQKGTGVAAGVGQW 679

RESULT 11  
US-08-617-697-10  
; Sequence 10, Application US/08617697  
; Patent No. 5977336  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/617,697  
; FILING DATE: 01-APR-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 05-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstresser, Jerry W  
; REGISTRATION NUMBER: 1038-557  
; REFERENCE/DOCKET NUMBER: 22,651  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1600 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-617-697-10

Query Match 5.08; Score 517.5; DB 2; Length 1600;  
Best Local Similarity 21.99; Pred. No. 1.5e-23;  
Matches 418; Conservative 239; Mismatches 766; Indels 487; Gaps 94;

Qy 1 MNHIYKVFNKATGTFMAVAEYAK--SHSTGGSGCATGQGVSVCTLSFARITAAALVLVIG 58  
Db 1 MNKYRLKFKSLNALVAVSRLTRGCDHSTKGEKPV-----TKVRHLAKPLS 51  
Qy 59 ATLSGSAVAOKKDKTKHAIQBNQPPRSGTA--RADGDRAITAIGENANA----- 105  
Db 52 AILLSGMASTPQSV-LASGLQGMSVVGHTATMQVDGKNT--TIRNSVNAIINWKQFNIDQ 109

Qy 106 -QGQATAIGASSNKTVNGSSLDKIG-----TDATQGESIAIGDGVKASGDASIAIGSDDL 159  
Db 110 NEMEQFLOESSNSAVFNRTSDQISQLKGLDLSNGQVFL-----INPNG---ITIGKD-- 159  
Qy 160 HLLDQHGPKPKGTGLINDLINGHAVLKEIRSSKDNQVYKRTTASGHASTAVGAMSAQ 219  
Db 160 -----AIINTNGFTASTLDSINIKARNFTLE---QTKKALAEIV 198  
Qy 220 GHFSNAFCTRATAKSAYSGLAVLAATAEGQSTIAIGSDATSSSLGAIALGAGCTRAQLQGS 279  
Db 199 NH-----GL-----ITVGKGSYNLIGGKYNKNEGVISVNGS 230  
Qy 280 IALGOGSVVVTOSDNNRPAYPNTQALDPKFOATN-----NTKAGPLSTGSSISKRKIINV 335  
Db 231 ISLLAGQKITISD-IINPTIYVIAA--PENEALNGLDIFAKGGINVRAATIRNKGKLS 287  
Qy 336 GAGYNKTTDAVNV-----AQLEAVV---KWAKERITFGQDDNSDVKIGLDTLTI 383  
Db 288 ADSVSKDGSNGNIVLSAKEGEAEIGVISAQNOQAKGGLMITGD--KYTLKTGAVIDLSG 345  
Qy 384 KGAETNALTDNNIGVVKKEADNSGLKVKLAKTLNLTETVNTTTLNATTVKVGSSSSTA 443  
Db 346 KEGGETYLGDER---GEGKNG---IQAK-----KTTLEKGSTINVSGKEGGR 389  
Qy 444 ELLSDSLTFTQPTNGSQSTSKTVYGVNGVKTNNAAETAAIGTTRITRDKIGFARDGVD 503  
Db 390 AIVMGDIALIDGNINAQSDIAKTG-----GFTVETSHDLSIGDVIAD 433  
Qy 504 EKQAPYLDKKOLKVGCSVAITIDN-----GIDAG-----NKKISNLAKSSANDAVTIQOL 553  
Db 434 AKEW-LDDPDVDSIETLTSGRNTGENOGYTTGDTKESPKGNSISKPLTNS--TLEQI 490  
Qy 554 KAAKPTLNAGAGISVTP-----TEISVDAKSONVTAPIYNICVKTTELSNDSGTSDK--F 605  
Db 491 -----LRGGSYVNITANNRIYVNSSINLSNGLTLHTRKRGVK---INGDITSNENGL 541  
Qy 606 SVKSGS-----TNNSLVTAEHLASLYNEVNRVTASALQSFTVKEEDDDANA----- 652  
Db 542 TIKAGSWVDVHKNTLGT-----GFLNIV-----AGDSVAFEREGDKARNATDAQITAQ 590  
Qy 653 --ITVAKD-----TTNAGAVSILKLGKNGNLVATKK 683  
Db 591 GTITVKKDDQFRFNNVSLNGTGKLGKFIANONNFTKFDGEINI---SGIVTINQTTKK 647  
Qy 684 DGTVTFGLSQDSGLTIGKSTLNNDLGTVKDTNEQIQVGANGIKFTNV-----NGSNPGTGI 739  
Db 648 D-VKYWNASKDSYWNVSSLTLN---TVQ-----KFTFIKFDVDSGSN-GQDL 688  
Qy 740 ANTARITRDKIGFAGSDGAVDTN-----KPYLDQDKLQ-----VGNVYKIINTGIN 784  
Db 689 RSSRR-SPAGVHFNGIGGKTNFNIGANAKALFKLPNAATDPKELPITFANITATG-N 746  
Qy 785 AGKRAITGLSPTLPSIA---DQSSRNIELG---NTIQDKDSNAASINDILN---TGNL 835  
Db 747 SDSSVMFDIHANLTSRAAGINMDSINITGGDLDFSTSHRNSNAFEIKKDLTINATGSNF 806  
Qy 836 KNNNPIDFVSTYD---IVDFANGNATTATVTHDTFANKTSKVVDVNVYD---DTTHLTG 889  
Db 807 SLKQTKDSFYNEYSKHAINSSHNLTLGNGVNLGGENSSSITGINITNKRANVTLOADT 866  
Qy 890 TDDNKKLGQVTKTKLNKTSANGN-TATNPNVSSDEBALVNAKDIAENLNTLAKEIHTTKG 948  
Db 867 SNSNTGLKKRLTLTGNISVEGNLSLTGANAN-----IVGNLSIAED-----STFKG 912  
Qy 949 TADTALO---TFTVKKVDENNADDAN---AITVCQKANNQOVTLTLKGENGLNIKTDK 1002  
Db 913 EASDNLNITGFT-----NNGTANINIKGVYKLGIDN-----NKGGLNITNA 955  
Qy 1003 NGTVFTGLNTTSGLKAGKSTLNDGGLSLKNTPGSQIQVGADGVKFAKVNNGNVVVGAGID 1062  
Db 956 SGTOKTIIN-----GNITNEKGLNINIKKADAIEIQIGN-----ISQK-----E 995

Qy	1063	GTTTRIDREIGFTVTCNSLDKSPHLXSKDGINAGCKKITNIQSCEIAQNASHDAVTGKIY	1122
Dy	996	NLTISDDKNVIT-----NQITIKAGEGRSDSEAEANLTIQTKEKLAKG----	1043
Qy	1123	DLKTELENKISSAKTAQNSLHEFSVADEOQNFTVSNPYSYDTSTKTSDVITFAGEN--	1180
Dy	1044	DLNISGFNKAEITAKNGS----DLTIGHASGGNADAKK--VFDPKVDSKLST-DGHVNT	1096
Qy	1181	-GITTKVNKGWVRVGIDOTKGLT--TPKLITGVNNNGKVIDSGONGQNTITGLSNTLANV	1237
Dy	1097	LNSEVKTSSNGSNAGDNSTGLTISAKDVTNVNRVTSHTKTINISAAAAGNVTTKEGTTINA	1156
Qy	1238	TNDKGSVRTTEOGNIILKDEDKTRAASIVDVLGSAFNLOGNEAVDFYSTVDTVNFPADGNA	1297
Dy	1157	TT--GSSEVTAQNGTIK-----GN-----ITSQNVTVTATENLV	1188
Qy	1298	TTAKRVTYDDTSKTSKVYDVNWDDTTIEVKDKLKGVKTTT---LTSTGTGANKFAALSNO	1353
Dy	1189	TTENAVINATSQT-----VNISTKTGDIMG--GIESTSGNVNITASG---NT'LKVSN-	1235
Qy	1354	ATGDALVKASDIVAHNLTLSGDIQTAKG-----ASQAANSAGYVDADGNKVLYDSTDNKYY	1409
Dy	1236	ITGODVTVTADAGALTATTAGTISATTNGANITTTTGTDINGKSVSSGSVTLVATGATLA	1295
Qy	1410	QAKNDG-TVPDKTEVAKDKLVAQ-AQTPDCTLAOMNVKSVINKEQVNDANKKOGINEDNA	1467
Dy	1296	VGNISGNTVITIAD--SGKLTSTVGSTINGT-----NSVTTSQSGD-----	1335
Qy	1468	FVKGLEKAASDNKTKNAAVTVVDLNNAVAOTPLTFPAGDTGTAKKLGETFLTIIKGQTOTNK	1527
Dy	1336	----IEGTISGN-TVNVTASTGD-----LTICNSAKVEAKNGAATLTAESG-----K	1377
Qy	1528	LTDNNIGVWAGTGTGFKLAKDLTLNLSYNAGGRTKIDBKGSFVDSGSOAKANTPVLSAN	1587
Dy	1378	LTTOGSSIYSSNGQTTLAKDSSIAGNINAANVTLNTIGLITTTGDSKINATSGITLIN	1437
Qy	1588	GLDL-----GKVISNV-----GKG-----TKDPTDAANVO-QLNEVRNLLGLGNAGNDNA-	1631
Dy	1438	AKDALDGAAAGSORTVVYVATNATNASGSGNVTAKTSSSVNITGDINTINGLNIISENGRVTJR	1497
Qy	1632	-DGOVNATADIKPDNPGSSSSNRTVIKAGTVL---GGKGNNDTEKLATGGIQ-VGVPDKG	1686
Dy	1498	LRGREIDVKYIQ-----PGVASVEEVLKARVLEKVKDKLDSDERETAKLGVSANRVFEPN	1553
Qy	1687	NANGDLSNVWKYTKDQGSKKALLATYNAAGQNTNYLTNNPAEAIDRINEQG	1736
Dy	1554	NA-----ITVNTONEFTTKPSVOSITSEGKACFSGGNGARCVCTMVWADG	1597

RESULT 12

US-08-617-697-9

03 00 017 037 2  
; Sequence 9, Application US/08617697

; Patent No. 5977336

: FACILE NO. 5577530  
 : GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J.

APPLICANT: Bartenkamp, Stephen S.  
TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE POLYMERIZATION OF VINYL MONOMERS

NUMBER OF SEQUENCES: 11

; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: shoemaker and Mattare. Ltd.

ADDRESSEE: Shoemaker and Mattare, LLC.  
STREET: 2007 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: 2001 Je  
STREET: Bldg 1

STREET: Bldg. I  
CITY: Arlington

CITY: Arlington  
STATE: Virginia

STATE: Virginia  
COUNTRY: U.S.A.

COUNTRY: U.S.A.  
 ZIP: 22202-0796

; ZIP: 22202-0286  
; COMPUTER READABLE FORM:

COMPUTER READABLE FORM:  
MEDIUM (MYOE): Floppy disk

```

;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;

```

COMPUTER: IBM PC compatible

```

;
OPERATING SYSTEM: PC-DOS
COMMENTS: Detectable Rel

```

```

; SOFTWARE: PatentIn Release #1.0,
; CURRENT EDITION: 1.0

```

Db 606 VSLNSLGKLSFTDSREDRGRRTKG--NISNKFDTGLNISCTVDSMKAPKVSFWRDKG 663  
Qy 1004 CT--VTEGINTTSLKAGKSTLNDG---GLSIKNPTGSEIQVGADGVKFAKVNNGV 1056  
Db 664 RTYNNVT--TLNVTSGSKFNLSDTSGSGTSPISRN-----AELNGITFNKATFN-- 712  
Qy 1057 VGAGIDGTRITREIGF--TGTNLSLKSPLSKDG-----INAGGKKITNIOG--GEIA 1109  
Db 713 IAQSTANFISAKISIMPEKSNANTALEFNEDISVGGSVNFKLNASS---SNIQTGPVII 769  
Qy 1110 QNSHDVATGGKIYDLKTELENKISSTAKTAQNSLHEFSVADEQGNFTVSNPYSSYDTSK 1169  
Db 770 KSONFNWGGTLLNLAB-----GSTETA-----FSIENDLNUNATGNN----- 808  
Qy 1170 TSDVITFAGENGITTKVNGVVRVIGIDQTKGLTTPKLTGVNNG---KGIVIDSQNGQN 1225  
Db 809 ----ITIRQVEGTSRVNKGVAANKNITFG---GNITFGSOKATTEIKGNVTINKNTNA 861  
Qy 1226 TITGLSNFLANVTNDKGSVRTEQGNITIKDEKTRAASIVD-----LSAFENLQG--NCE 1279  
Db 862 TLRG-----ANFAENKSLNTA--GNVINGNLTAGSIINAGNLTVSKGANLOAITNY 914  
Qy 1280 AVDFVSTYD-----TVNFDAGNATTAKVTYDDTSTKSVVYDVNVDPTIEVKDKLGVK 1334  
Db 915 TPNVAGSFDNNGASNISLARGA-----KPKDINNTSSLNITVNSDITYRVI----- 961  
Qy 1335 TTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSSGDIQAKGASQANNSAGYVDA 1394  
Db 962 -----IKGNISNKSGLDNIITDKKSDAEIQIG---- 987  
Qy 1395 DGNKVIYDSTDNKYIOAKNDGTVDKTEKVKDKLVQAQOTPDGTLAQMNVKSVINKEOVN 1454  
Db 988 -----GNISREGNLTITSSDRVN 1005  
Qy 1455 DANKQGINEDNAFVKGLEKAAD--NKTNAATVVDGLNVAQOT--PLTFAGDTGTAKKL 1512  
Db 1006 -----ITNQITKAGVEGRSSSEANLTI-----QTKELAGLDNLISGNK 1051  
Qy 1513 GETLTKGGQDRTNKLTDNNIGVAVAGTGFVKLAKDLTLNLSVSNAGGTKIDDKGVSEVD 1572  
Db 1052 AE--ITAKNG-----SOLITIG-----NASGKNADAKKVF--D 1080  
Qy 1573 SSGQAKANTPVLNGLDLGKVISNCKGKTGTDAAVVOQLNEVRNLLGLGNAGNDNAD 1632  
Db 1081 KVKOSKIS-----TDGHNVTLSNLTSEVKTSGSSNAGNDNST 1115  
Qy 1633 GNOVNIADIKDPNPGSSSNRTV---IKAGTVLGKGNNDTEKLTATGIVGVVDKDGAN 1689  
Db 1116 GLTISAKDV--TVNNVTSHKTIINISAAAGNVTTKEGI--TINATGSGVEV--TAONGTIK 1170  
Qy 1690 GDLSNVWYKQKDGSKKALLATYNAAGQTNLTNNPABEIDRINEQGIFFHVNNDGNOEP 1749  
Db 1171 GNITS-----QNVTVTATENLVTTENAVINATSGTVNISTKTG----- 1208  
Qy 1750 VVOGRNGIDSSASCKHSAVGAFOAKDEAAVAGRTQAGNQSIAGDNAAQ--TGDOSI 1808  
Db 1209 ---DIKGGIESTSGNNITASGNTLKVSNITQDVTVTADAGALTTTAGSTISATGNANI 1266  
Qy 1809 AIGTGNVYAGK---HSGAI-----GDPSTVKAADN-----SYSGVNNN 1842  
Db 1267 TTKTGD--TKNGVSSGSVTLVATCATLAVGNISGNTVTIITADSGKLTSTVGSTINGTNS 1325  
Qy 1843 QFTDATQDVFVG--GNNITVYES--NSVALGSNSAISAGTHAGTOAKKSDGTAGTTTAT 1899  
Db 1326 VTTSSQSGDIECTISGNTVNVNTASPGDLTIGNSAKVEAKNGAATLTAES--CKLTJTQTGSS 1384  
Qy 1900 ATGTVKGFAGQATVAGCAVSGASGAERRIQNVAAAGVYSATST-----DAVNGSQL 1948  
Db 1385 ITSS-----NGQTTLTKADSSSTAG-----NINAANVTLLTGTTLTTGDSKINATSGT 1434  
Qy 1949 YKATQSIANATNELDHRTHQENKFNKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGOG 2008

Db 1435 INAKDAKLDGAASGDRTVVVNATNAGSGNVTAKTSSSV-----NITGDLNTIN--- 1482  
Qy 2009 AVAVGSLKSLSDNGQWVFKING 2029  
Db 1483 ---GLNIISENGRNTVRLRG 1499  
RESULT 13  
US-08-728-470-10  
; Sequence 10, Application US/08728470  
; Patent No. 5928651  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728,470  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-633  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1529 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-728-470-10  
Query Match 4.9%; Score 501.5; DB 2; Length 1529;  
Best Local Similarity 21.3%; Pred. No. 1.3e-22;  
Matches 373; Conservative 238; Mismatches 670; Indels 471; Gaps 85;  
Qy 425 TTLNATTVKVSSSTTAELLSDSLTTQTPNTGSGSTKTVYGVNGVKFTNNAETAAI 484  
Db 3 SVVHGTTATQVDGNKTTIRNSVNAIINNKQFIDQNE-----QFLOESSNSAVF 53  
Qy 485 GTTRITRDKIGFARDGVDEK--QAPYLDKKQLKVGSAVITDNGIDACKNKKISLNLAGSS 543  
Db 54 --NRVTSQISQLK--GILDSNGQVFLPNPNTIGKDAIINTNGTASTLDISN--ENIK 108  
Qy 544 ANDAVTIEQL--KAAKPTLNAG-----AGISVTPTEISVDAKS-- 579  
Db 109 ARN--FTLEQTRDKALAEIVNHLITVKGDSVNLIGGKVKNEGVISVNGGSIILLAGQKI 167

QY 580 --GNVTAPTNYGVKTTLSNDSGTOKFSVKGSGTNNLSVTAHLASYLNEVARTADSAL 637  
 Db 168 TISDIINPTITYSIAAPENAINGLDIFA-KGGNIN-----VRAATIRNKGLSADS- 219  
 QY 638 QSFTVKEEDDDANATVAKD-TTKNAGAVSILKLKGNKLTVA- 681  
 Db 220 -----SKDKSNIIVLSAKEGEAGCGVISAQNOQAKGKLMITGDKVTLTKGAVIDL 272  
 QY 682 KDKGTWTFGLSQD-----SGLTIGKSTLNMNGLTVKDTNEQIQVGANGIKFTNV----- 730  
 Db 273 GKEGGETY-LGGDERGEGNGIQLAKKTTLEKSGTINVSCK--EKGGRAIVWMDIALIDG 329  
 QY 731 NGSNPGTGANTARIYRDKIGFA-----GSDGAVDTNKPYPLODKLQGVNWKI--T 779  
 Db 330 NINAQSDIAKTG-----GFVETSGHDLISIGDDVIVDAKEWLLDPDDVSIETLTSGRN 382  
 QY 780 NTGINAGGRAITGLSPTLPSIADQSSKNIELGNTIQDKDKSNAASINDILANTG--ENLKN 837  
 Db 383 NTGENOG--YTTG-----DCHKESPKGNSISKPLTIN-STLEQLIRGSGYVNITA 429  
 QY 838 NNPNIDFVSYDIDVDFANGNATTATVHTDANKTSKVWYDVNVDDTTIHLTGDDNKKLG 897  
 Db 430 NNR-----IYVNSSINLSGSLTL-----HTKRDGVKINGDIT-----SNENGLT 470  
 QY 898 VKTTKLKNTSANGTATNP-NVNSDEEDALVNAKDIAENLTLAKELHTTKGTADTALQT 956  
 Db 471 IKAGSVWDVHKNTLTGTGLNIVAGDSVAFEREGDKARN-----ATDAQITAQGT 520  
 QY 957 FTWKVKYDEN--NNAADANAITVGOKNANNVNLTLLKGENGLINKTKNGVTTFPOINT 1013  
 Db 521 IYVKKDKQFRFNV-SINGTKGLKFAIANO-NFTHKFDGELNI-----SGVITINQTKR 574  
 QY 1014 SGLK-----AGK-STLNDGGLSIKN-----PTGSEQIQVGADGVKFAKVNNGVGA 1059  
 Db 575 KVKYVNAKSDYVWVSSLTNTVQKFTFKFVDSGSSNSQDLRSSRRSPAGVHFNIGCK 634  
 QY 1060 GIDCTTRITRDEIGFTGNSGLDKSPHLSKDGINAGGKKTNIQSGEIAQNSHDVATGG 1119  
 Db 635 -----TNFNIGANAK--AFUFLKPNATD-----PKK-----ELPITFNANITATGN 674  
 QY 1120 K-----IYDLKTELENKISSTAKTAQNLHEFVSADQGNFTVSNPYSSYDTSKTSDDVT 1175  
 Db 675 SDSSVNFEDIHANL-----TSRAAGINMDSINITGLDFSIHSHRNSNAFELKDLTLIN 728  
 QY 1176 FAGENGITTKVNGGVVRVIGIDQTKGLTTPKLTGVNNGKGVIVDSQN-----QONTITGLS 1231  
 Db 729 ATGSN-----FSLKQTKD-----SFYNEYSKHAINSSHNLTLILGQNVTLGGE 770  
 QY 1232 NTLANVTNDKGSVRTTEQNIILKDED-----KTRAASIVDVLVSAGFNLOGNGEAVDF 1283  
 Db 771 NSSSIT-----GNINITNKANVTLOADTSLNSNTGLKRTLTGLNISVEG-NLSLTGANANI 826  
 QY 1284 VSTYDVTNFDAGNATKAVTYDDTSKTSKVWYDVNVDDTTIEVKDKKLGKVTTLTSTGT 1343  
 Db 827 V-----GNLSTA-----EDSTFK-----CEASDNLNTGT 851  
 QY 1344 GANKFALSNOATGDALVKAASDVAHLNLTLSGDIQTAKGASQANNSAGYVDADGNKVYDS 1403  
 Db 852 -----FTNNGTANINIKGVV-----KLQGDINKNGGKNITTNASG-----TQKTLIN 896  
 QY 1404 TONKYVQAKNDGTVDKTEKAVDKLVAQAQTPDGTLAQNNVKSVINKEOVNDANKKOGIN 1463  
 Db 897 ITNE-----KGLNLIKADAEIQI-----GGNISQKEGMLTSSDKVN-----IT 938  
 QY 1464 EDNAFVKGLEKAASD-NKTKNAAVTVGDLNVAQAQ--PLTFAGDTGTAKKLGELTLTKG 1521  
 Db 939 NQITIKAGVEGGRSDSSEANANLTI-----QTKELLAGDLNLSGNPKAE-ITAKNG 990  
 QY 1522 QDTHKLTNNNVTGAGTGTGVKLAKDLTLNLSVNAAGTKIDDKGVSPVDSGQAKANT 1581  
 Db 991 -----SDLTIG-----NASGKNADAKKVTF-DRVKDSKIS- 1019

QY 1582 PVLISANGLDLGGKVISNVGKGTDKDQDAANVQOLNEVRNLLGLGNAGNDNADGNNOVNADI 1641  
 Db 1020 -----TDCHNVTLNSEVKTSNGSSNAGNDNSTGLTISAKDV 1055  
 QY 1642 KKDPSNGSSNRTV---IKAGTVLGGKGNNDTEKLTAGGIVQGVDDKDNANGDLSNVVVK 1698  
 Db 1056 --TVNNVTSHTKINISAAAGNVTTKEGT--TINATGTSVEV-TAONGTIKGNITS----- 1106  
 QY 1699 TOKDGSKALLATYNAAGOTNYLTNNPAEAIIDRINEQGRFFHFHVDNGNOEPVQVQGRNGID 1758  
 Db 1107 -----QNVTVTATENLVTTENAVINATSGTVNLSITKG-----DIKGGIE 1146  
 QY 1759 SSASGKHSVAIGFOAKADGEAAVAIAQRQQAQNOISAIQNAQA--TGDQSIATGTGNVA 1817  
 Db 1147 STSGNVIITASGNITLKVSNITGQDVTVTADAGALTITTAGSTISATTGNANITTKGD-IN 1205  
 QY 1818 GK---HSGAI-----GDPSTVKADN-----SYSVGNNOQTDTATQTD 1851  
 Db 1206 GKVESGSGSVTLVATGATLAVGNISGNVTVITADSKLSTVSGSTINGNSVTTSSQSGD 1265  
 QY 1852 VFGV--GNNTITVES--NSVALGSNSAISAGTHAGTQAKKSDGTAGTITTAGATGTVKGFA 1908  
 Db 1266 IEGTISGTVNVTASTGDLTIGNSAKVEAKGAATLTAES-GKLTTQTGSSITSS-----N 1320  
 QY 1909 GQFAGVAGSVGASGABERRIQNVAAGEVSATST-----DAVNGSOLYKATQSIAN 1957  
 Db 1321 GOTTITAKDSSIAG-----NINAANVTLTNTGTLTTGDSKINATSGTLTINAKDAKLD 1374  
 QY 1958 ATNELDHRITQENKANAGISSAMAMASMPQVIPGRSVMTGGIATHNGOGAVAVGLSKL 2017  
 Db 1375 GAASGDRTVVNATNAGSGNVTAKTSSSV-----NITGDLNTIN-----GLNII 1418  
 QY 2018 SDNGQWVFVKING 2029  
 Db 1419 SENGRTNVTBLRG 1430

RESULT 14  
 US-08-719-641-10  
 ; Sequence 10, Application US/08719641  
 ; Patent No. 6218141  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barekamp, Stephen J  
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
 ; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Shoemaker and Mattare, Ltd  
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 ; STREET: Bldg. 1  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202-0286  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/719,641  
 ; FILING DATE:  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/302,832  
 ; FILING DATE: 16-SEP-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US PCT/US93/02166  
 ; FILING DATE: 16-MAR-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GB 9205704.1  
 ; FILING DATE: 16-MAR-1992  
 ; ATTORNEY/AGENT INFORMATION:







Db 1181 KLDGAASGDRTVVWATNAGSGNVTAKTSSV-----NITGDLNTIN-----GL 1224  
Qy 2015 SKLSDNGOWFKING 2029  
Db 1225 NIISENGRNTVRLRG 1239

Search completed: September 13, 2001, 12:37:49  
Job time: 48185 sec



A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; F.R.J.D.; Junqueira, M.B.; Kemper, E.L.; Klicajina, J.P.; Krueger, J.E.; Kuramae, E.E.; La Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins Ay,A; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C., F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, Rodrigues, V.; Rosa, A.J.de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawada,Autroques, da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da SilvM.: Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vetture, A.L.

A:Reference number: AS9328

A:Contents: annotation

C:Genetics:

A:Gene: XF1529

Query Match                11.1%; Score 1143; DB 2; Length 2059;  
Best Local Similarity     24.0%; Pred.No.1.4e-36;  
Matches 547; Conservative 342; Mismatches 766; Indels 626; Gaps 104;

QY      86 SGTA--KADGDAIATIGANAGGAIAIGSNKT--VNGSSLDKITGDATGOESIAIG 141  
||||| :|| : || | ||| : ||| : ||| : ||| : ||| :  
Db          86 SGTAASQGQSARNLTLGGSLVVNSGVGVNDVLNKITYIRMGSVITMTNTVA-GTNAIAIG 144

QY      142 GDVKASGSADIATGSDDLHLDDHQGNPKHPKGPTLINDLINGHAVLKBRSKDNKVYRR 201  
:: || | : |:  
Db          145 SAOSSAADAL-----KASL-----A 159

QY      202 TTASTGHASTAVGMASYAQHFSNAFCTPATAKSAYSILAVCLAATAECOSTIATCSDATSS 261  
| || || : || | :: | : | : ||| : ||| : ||| :  
Db          160 TRASGARFATIGAkakSGADVDTVALGSGATA-----GTGCASSIATGLNASAV 206

QY      262 SLGAIALGAGTRAOQL-OGSIALGCGSVVTVOsdNNRSPAYTPNTQAOLDPFOQTNNTKAGS 320







Db 750 ----IVLSKESGSN-----||| 762  
Qy 1271 GFNLGNGEAVDFVSTYDVFADGNATTAKVYDDTSKTSKVYVDVNDVDTTIEVKDKK 1330  
Db 763 -FNLSSS-----||| 773  
Qy 1331 LGVKTTLTSTGTGANKFALSNQATGDALVKASDIVAHLNLTSGDIQTAKGASQANNSAG 1390  
Db 774 LTVGDTVMVTNG-----||| 785  
Qy 1391 VVDADGNKVIYDSTDNKYYQAKNDCTVDKTRVAKDKLVQAQTPDGTQLAQMNVKSVINK 1450  
Db 786 -----||| 785  
Qy 1451 EQVNDANKKOGINEDNAPVKLEKAASDNKTNAAVTVGDLNVAQAUTPLTFAGDTGTTAK 1510  
Db 786 -----VTVG-----SGVTLG 795  
Qy 1511 KLGETLTKGGQDTNKLTDNNIGVVGAGTGTFTVKLAKDLNLNSVNAAGTKIDDKGVSF 1570  
Db 796 SMGLVI-----TDG-----804  
Qy 1571 VDSSQAKANTPVLISANGLDLGGKVISNVGKCTKTDAAVNQOOLNEVRNLLGLGNAGNDN 1630  
Db 805 -----PSVTSSGINAGSKITNVAAGTADTDVNLSQLN-----838  
Qy 1631 ADGNOVNIADIKDPNSGSSSRNRTVIKAGTVLGGKGNNDTEKLAT--GGIQGVDBDKGNA 1688  
Db 839 -----TAMAGSAGKSVHYSTYDGGTQ-----GGNY 864  
Qy 1689 NGDLSNVWVKQKDGSKKALLATYNAAQTNYLTNNPAEADIRINEQIRFFHVDNGNOE 1748  
Db 865 NGD-----867  
Qy 1749 PVVQGRNGIDSSAGSKHSAVFQAKADGEAAVAIGRTQA--GNQSIAGDNAAQTGDS 1807  
Db 868 -----GAYGTRIAVGVGTSLASAGATAVGSAAASGKSTAGRNVAASAGS 916  
Qy 1808 TAIGTG-----NVVAKHSG-----AIGDPSTVKADNSYSVGNNNQFTDAT--1848  
Db 917 VALDGAKDGARGAESYTGKYSGLQNNVTGTVSGDASKGETRVTNSVNDAKEATDAVNL 976  
Qy 1849 -QTDVFG-----VGNNI-----TVTESNSVALGSNSA--ISAGTHA-----GTQAKK 1887  
Db 977 RQLDRVAQDANRYVDNKTESLSEGTFTVKVNSL-----NNSATPIAAGVDATAIGVGATASG 1033  
Qy 1888 SDGTA-GTTTATAGTGVKGFAGQTAVG-----AVSVGASCAERRIONVAAGEVSA 1937  
Db 1034 ADSIANGKASASADNAV-----AIGNHVSADRANTVSVGSAGSERQVNTNVAAG----T 1083  
Qy 1938 TSTDVANGSOLYKATQSIANATNELDRIHONENKANAGISSAMAMAMPQAYIPGRSMV 1997  
Db 1084 ADTDVAVNSQL--NOGLITAKQYTDGVVGSILRRDTDGVAATAATNLQAYIPGRGMT 1140  
Qy 1998 TGGIATHNGQCAVAVGLSKLSDNGOWPFKINGSADTQGHVGAAGVAGPHF 2047  
Db 1141 SVGVSSYRQCAIAVGVSSVSESGRWFKFGCSANTRSQVGTGAGVGYQW 1190

RESULT 4  
B85547  
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: B85547  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: B85547  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-5188 <STO>  
A:Cross-references: GB:AB005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:  
C:Genetics:  
A:Gene: Z0615

Query Match 6.2%; Score 641; DB 2; Length 5188;  
Best Local Similarity 22.0%; Pred. No. 1.le-16;  
Matches 519; Conservative 297; Mismatches 872; Indels 668; Gaps 113;

Qy 34 ATCGVGSVCTLSFARIAALAVLIVIGATLSSGSAAYAKOKTKHIAIGEQNPPRRSGTAKADG 93  
Db 1436 ATAPSVTINTTADTDLNAAEAGSALTISGTSTAEAGOTVTVTLNGVNY---SGNVQADG 1492  
Qy 94 DRAITAIGENANAQGOATAIGSSNKTVNGSLDKITGTDATGOESIAIGDVKASGDASTA 153  
Db 1493 SWSVS-----PTGDLASLTASSVTVNASVSDKARNASATHNLTV--DLAAPVVTINT 1544  
Qy 154 IGSDDLHLLDQHGNPKPGKTLINDLINGHVLKKEIRSKDNDVKYRRTTASCHASTAVG 213  
Db 1545 VAGDDIINATHEGQAO-----IISGAT-----GATTGNTVSVTIG 1580  
Qy 214 AMSY-----AOGHFNAGFTRATAKSAYSLAVGLAATAEQOSTI-----ATGSDATSS 262  
Db 1581 TTTYTTVLVDANGNWS--ICVPASV-----ISALAQGDVTITATVTDASGNSGTASH 1629  
Qy 263 LGAIALGAGTAAQLOGSI-----ALGQSVVTVQSDNNSRPAYTPNTQALDPKFOATNNTK 317  
Db 1630 TVTVGALGAPVLAINITIAVDDIINAAEKADLAITGTSNOPAGQTIVTVLNGQNYTTTADA 1689  
Qy 318 AGPLSI-----GNSIKRKIIINVAGVKNKTDVAVNAQ 349  
Db 1690 SGNWSVTVPASRVSAIGEATVYTAATDADGNSGSASHNVQVNTALPGV----TINVVA 1745  
Qy 350 LEAVVYKAK---ERRITFGDDNDSDVKIGLDNTLTIKGAETNALTDNINIVKVEADS 406  
Db 1746 TDDIINAAEAGVEQTIISGQ---VTGAAGADTVTVTL-CGATYTTATVOANLSWSDVPAS 1800  
Qy 407 GLKVKLAKTLNLTETVNTTLNATTVKVGSSSTTAELLS-----448  
Db 1801 AL-----GELGN-----GELTISASVNSVNGTNGTRETITIDANPLGLRVDTVAGDDVNV 1851  
Qy 449 -----SITFQPNTGQSOSTSKTVGVNGVKFTNN--AETTAAGTTRITRDKIGFARDG 500  
Db 1852 IIEHQALVITGSSSGLAAGSNVTLTINCQTVAAVLADGTWSVG-----1896  
Qy 501 DVDEKQAPYLDKKQKLVGSAITIDNGDAGNKKISNLAAGSSANDAVTIEOLKAAKPTL 560  
Db 1897 -----VPAVDVSAMPAGSVTIAASGSTSAGNP-----VSVTHPVTVD--LSAVAVSI 1941  
Qy 561 NA-----CAGISVTPTEISVDAKSGNVTAPTYNIGVKT-----593  
Db 1942 NAITADDVINAKEGAALTLSGSTSGVEA--GQTVTVTF--GGKTYSATVAANGSNSTSV 1997  
Qy 594 -----TELNSDGTSDKFSVSGTNNSLVTAELHASYLNEVNRATDSALQSTVVEEDD 648  
Db 1998 PAADMAALRDGDASAQASV--SNVNGNSATTHAYS-----VDASAPVTINTIAGD 2047  
Qy 649 D-----ANAITVAKDTTKNAGAVSILKKGKNGL--TVATKKDGTVTPEGLSODSGLTIG 700  
Db 2048 DILNAAEAGAALTITGSSSTAEAGQTVTVTLNGTNTGTGTGDSGVSVPASDLSTLTAS 2107  
Qy 701 KSTLN-----NDGLTVKDTNEQI-----QVCANGIKFTN-----VNGSNPG 736  
Db 2108 XYTVNAAVSDRAGNPNASVNHNLTV--DTSVPVVTINTVAGDDVDVINATEHAQAIIISGATG 2166  
Qy 737 TGIANTARITRDKIGIFA---GSDGAVDTNKPVLDDOKLQVGNVKITNTGICINAGGKAITG- 792  
Db 2167 AATGSTVTVTCTNTFTTVLVDASGNWSVGPASVVSALANGVTVINASVTDAGGNSGAT 2226  
Qy 793 ----LSPTLPSIADOSSRNIELGNTIQDKKSNAAASI---NDILNTGPNLKNNNPIDFV 845



```

Db 2227 HQVTNGLPTITFNA---ISDNILNADEKQPLTISGSGTCLATGAQVTVTLNGHNS 2283
QY 846 STYDIVDFANGN-----ATTATVTHDTANKTSKVYVDVNVDDTT 884
Db 2284 ATTD-----ASGNWTLVPVSDLAALGAQANYTVSASATSAAGNTASSOANLLVDSGLPDVT 2339
QY 885 IHLTGTHD---NKKLGKVT---KLNKTS-----NGNTAT-----NFNVSS 921
Db 2340 INTVAGDDIINAAEAGADQTSIGVTRAAAGDVTVTGLGNTYTATVQNSLSWSVSPFA 2399
QY 922 DEDALVNAKDIAENLAKETHTTKGTAD-----TALQTFYTKVKVDENNADANA- 973
Db 2400 DLQALNG-DLTTTASVYRWANGTSGTRDTITDANLPLGRVDTVAGDDIVNSIEHQAL 2458
QY 974 -ITVGOKNAN-NOVNTLTLKG-ENGLINIKDKNGTVTFGINTSGLKAKSTLN-DGGLS 1029
Db 2459 VITGSSGLNAGAVLTVTNSVAYSATVQADGWSVGIIPAANYSAWPAAGPLTVEVDGSS 2518
QY 1030 IKNPTGSEQ---IQVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFTCTNGSLDKSKP 1086
Db 2519 ANNPVSVSHPTVLDLTAVALISINTVASDDVINAEKGT----- 2556
QY 1087 HLSKDGINAGGKKTINIOSGEIAQNSHDAVT-GGKIY-----DLKT-- 1126
Db 2557 NLTLGSGTSG-----IESGQTV-----TVTFGKTYTASVAANGSWSVNPAADLATLP 2605
QY 1127 -----ELENKISSAKTQNSLHEFVSADGQNNFTVSNPYSSVD---TSKTSVITFAG- 1178
Db 2606 EGAANVOASVSASGNSASATHAYSV-DASAPTLTI-NTIASDDILNAAEAGSPLTISGT 2663
QY 1179 ---ENGITTKVN-KGVVRVGDIDQTKG---LTPKLTVGNNGKGVIVD---SONGOMTIT 1228
Db 2664 STAEITGVTVTLNGATYITGVQADGWSVSVPTISALGALNASNYTVSATVNDKAGNPGS 2723
QY 1229 GLSNTLANVNDKSGVTRTEGNIKIDEDTRAASIVDVILSAGFNLOGGEAVDFVS--- 1285
Db 2724 ASHNLAVDTTAPVLTINTVAGDDIINDAEHAQAL-VISGTSAG-----GEXGDVSVVL 2776
QY 1286 ---TYDVNFADGNATKATVYDDTSKTSKVVDVNVDDTTIEVKDKLGVKTTTLTSTG 1342
Db 2777 NGKTYTTLTLDASGNS-----VGVPAADVTALG 2804
QY 1343 TGA-----NKFALSNQATGDALVKASDIVAHLNLTSGD---IOTAKGASQANNSAGY 1391
Db 2805 SGAQITIASVSDRAGNSDDASRTVTVLSAPVISINTIAGDDVINATEKGSDLA--LSGT 2862
QY 1392 VDADGNKVIYDSTDNKYQAKNDGTVDKTEYAKDKLVQAQAPDGTGLAOMNVKSVINKE 1451
Db 2863 SDQAPGTAITVTLNGQNYSAITDASGNSVTVP-----ASAVSALGATYSVTSVNTN-- 2915
QY 1452 QVNDANKKQGINEDNAFYKLEKASDNKTKNAAVTVGDLNVAQAQTPLTIFAGDTGTTAKK 1511
Db 2916 -----AQHNSSTASHNVQNTALPGITINPVATDDIINASEAGSAQTI 2958
QY 1512 LGETLITKGGQDTNKLTDNNIGVAGVDTGTVKLAKDLTNLNSVNAAGTKIDDKGVSVFV 1571
Db 2959 SGQVTGAAAGSTVTVEL-----GGKTYTATVQADLSLWNSVPAADWQ-----A 3001
QY 1572 DSSGOAKANTPVLSANGLDLGGKVISNVGKTKD--TDAANVOQLNEVRNLLGL---GNA 1626
Db 3002 LONGELTVNASVTNA-----VGNVTSRTRITIDA-----SLPGLRVDVTA 3042
QY 1627 GNDNADGNQVNIAD- IKKDPNPGSSSNRTVITKAGTVLGGKNNDT-----EKLATGGIQGV 1682
Db 3043 GDD-----VNIIEHAQAVQVITGSSG---FAAGTALTIVVNNQTYAATVLANGSWSVGV 3094
QY 1683 KDXGNAGDLSNVWTKODGSKKALLATYNAAGQTNVLTN-----NPAEAI 1729
Db 3095 PAT-----DVSN-W-----PAGTLNVTSGANSAGTQTSITHTPLVDTLTAVALSINSTSD 3144
QY 1730 DRIN--EQCIRP-----FHVNDGNQEPVVOGRNGIDS--SASGKHSV---AIGFOAKAD 1776

```

```

Db 3145 DAINAAERGAALTISGSGTSGVEAGQTVTVTFEGKTYTTTVAANGSWSTVPAADLAALRD 3204
QY 1777 GEAAVAGIROQAGNOSTAIGDNAQATGDQSI-----AIGTGNVAGKHSGA--- 1823
Db 3205 GDASAQV-RVTNVN-----GNSATATHEYSVDSNAPVTINTIASDNIINASEAAGVT 3257
QY 1824 ICDSTVTKADNSYSV---GNNNQFTDAPQTDVFGVGN-NITVTSNSVALGSNSAISAGT 1879
Db 3258 VSGTSTAQTGOTLTTLTNGTNYQTT--VQTD-----GSMSTLPASDLTALANNGYTLTAT 3311
QY 1880 ---HAGTOAKKSDG-TAGTIT-----TTAG-----ATGTVKG--- 1906
Db 3312 VSDLAGNLGSGASKGVTVDTTAPVISFNTVAGDDVINVEHQAQIISLTATGAVAGDRLV 3371
QY 1907 --FAGQ-----TAVGAVSVGASGAERRIQNVAAGEV--SATSTDAVNGS-----QLY 1949
Db 3372 VTIAQQOQVYVSTSDASGNSVGVPAVSV--ISGLADGTVTISATITDTSAGNSSTQTTHNVQVN 3429
QY 1950 KATQSIANATNELDHRHONENKANAGISSAMAMAPQAYIPGRSMWTTGGIAT--HNGQ 2007
Db 3430 TAAVSLSVSTISGDNLI---NAAEAG--SALTLSG-----TGTNFAATGTVTVTVLLNGK 3477
QY 2008 GAVAVGLSKLSNGOW 2023
Db 3478 GYSAT-----IQNSGSW 3489

RESULT 5
A41477
190K surface antigen precursor - Rickettsia rickettsii
C:Species: Rickettsia rickettsii
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 08-Oct-1999
C:Accession: A41477
R:Anderson, B.E.; McDonald, G.A.; Jones, D.C.; Regnery, R.L.
Infect. Immun. 58, 2760-2769, 1990
A:Title: A protective protein antigen of Rickettsia rickettsii has tandemly repeated,
A:Reference number: A41477; MUID:90354033
A:Accession: A41477
A:Molecule type: DNA
A:Residues: 1-2249 <AND>
A:Cross-references: GB:M31227; NID:g152465; PIDN:AAA26380.1; PID:g152466
A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for resi
C:Keywords: surface antigen; tandem repeat
F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>

Query Match 6.1%; Score 627; DB 2; Length 2249;
Best Local Similarity 22.2%; Pred. No. 1.3e-16;
Matches 477; Conservative 233; Mismatches 754; Indels 684; Gaps 108;

QY 97 IATGENA-----NAQGGQAIAGSSNKTGVNGSSLDKIGTDATGQE---STAIGGDVKA 146
Db 44 IATNNNAAFSNVYNNNNNNNEITAGVAN-----GTPAGGQNNWAFYGGDYTV 92
QY 147 SGDASTAIGSDDLHLDDQHGKPKPKG-----TLINDLI-NGHAVLKEIRSKONDVKY 199
Db 93 TADAADRI-----IKAINVAGTT--PVGLNTQNTVGSIIITKGNLPLVTLNACKSLTLNG 146
QY 200 RRTTASGH-----ASTAVGAMSYAQHFSNAGTATKATASAYSLAVGLATAGQSTIAI 254
Db 147 NNAVAANHGFADPADNVTYGLNIALGANAALITQSAAPS-----KITLAGNIDGGGTLTV 202
QY 255 GSDA-----TSSSLGAIAGAGTRAOLQSGI-----ALGGQSVVTSQDNNRSPAYT 300
Db 203 KTDAAINGTIGNTNALATVNVGAGT-ATLGGAVIKATTKTLTNASVLTITNANA----- 256
QY 301 PNTQALDPKFOATNNTKAGPLSLG-----SNSIKRKITNVGAGVKNKTDVAVNVAQL 350
Db 257 VLTGAID---NTTGGDNVGVNLNLGALSQVTDGIGNTNSLATISVAG-----TATL 305
QY 351 EAVVWKAERLITFOGDDNSTDVKIGLDNLTITKGGAEETNALTDNNIGVYKADNSGLKV 410
Db 306 GGAVIKATTKTLT-----DAASAVK--FTNPVVVTGAIDNTGNANN--GIVTFYCNSTVTC 357

```

QY 411 KLAKTNNLTVNTTTLNATTTTKVQSS-----SSTTAELL--SDSLTTFQPTNGS 459  
Db 358 NVGNT-----NALTATVNGAGLLQVQGVVYKANTINLTNDASAVTETNPVVVT 405  
QY 460 QSTSKTVYGVNG-VKFTNNAETAAIGTPTTRITRDKIGFARGDGVDEKQAPYLDKKOLKVG 518  
Db 406 GADNTGNANNGIVTFTGNTSVTGDIGNTAL-----ATVNVG 443  
QY 519 SVAITIDNG-IDAGNKKISLAK-----GSSANDAVTIBQLKAAKPTLN 561  
Db 444 AGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQT 503  
QY 562 AGAGISVTPTEISVDKSGNVTAPTNYNIGVKTTELNSDGTSDKFS-----VKGS-----GT 612  
Db 504 GNIGNTNSLTIISVGA--GTATLGGAVIKATTTKLTDAASAVKFTPNPVVVTGAIDNTGNA 561  
QY 613 NNSLTAEBHLASVLYNEVNRRTDASALOSFTVKEEDDDDDANAITVAKDPTTKNAGAVSILK 672  
Db 562 NNGIVTFTGNTSVTGDIGNT--NSLTIISVAGTATLGGAV-IKATTTKLTNAASVLT 618  
QY 673 GKNGLTATKDKGT-----VTGLSQDSG-----LTIGKSTLNNDGLTVKD 713  
Db 619 NANAVALTGAIDNTTGGDNVGVNLNGLSQTGDIIGNTNSLTIISVAGTATLGGAVIKA 678  
QY 714 TNEQIOVGANGIKFTN---VNGSNPGTGIANTRITRDKIGFAGSDGAVDTNPKPYLDQDK 770  
Db 679 TTKKITNAVASAVKFTPNPVVVTGAIDSTGNANGIVT-----FTGNTSVTG-----723  
QY 771 LOVGNVKITNTGKNAGKATITGLSPTLPSIADOSSRIELGNTI---QDKDKNAASIND 827  
Db 724 -DIGNTALAT-VNMGAGTAT-----LGGAVIKATTTKLTNAASVLT 763  
QY 828 ILNTGFNLKNNNPIDFVSTYDIDVFANGNATTATVTHDTANKTSKVYDVNVDDTTIHL 887  
Db 764 LTNANAVL---TGAIDNTTGGDNVGVNLNGLSQTGDIIGNTNSLTIISVAG-----814  
QY 888 TGTDDNKKLGKVTTLNK-----TSANGN---TATPNVNSDEDALVNAK-----D 931  
Db 815 TALGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQTG 874  
QY 932 IAEMLNTLAK-----EHTTKGTADTALQFTVKVVDENNADANAI-- 974  
Db 875 IG-NTNSLTIISVAGTATLGGAVIKATTTKLTNAASVLTLT-----NANAVALTGAIDN 927  
QY 975 -----TVGQKANNQVNTLTKGNGLNITDKNGVTFTGINTTSLGKAGKSTINDGGLS 1029  
Db 928 TTGGDNVGVNLNGLSQT--GDIG---NTNSLTIISVAG-----ACTATL--CGAV 972  
QY 1030 IKNPTGSEQIOVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFTGNT--GSLDKSKP 1086  
Db 973 IKATT--TKLTDAAAVKF---TNPVVVVTGAIDNTGNANGIVTFTGNTSVTGNVGTNA 1027  
QY 1087 HLSKDGINAGGKKITNTQSGEIAQNSHDAVTGGKIYDLKTELENKISSAKTAQNSLHEF 1146  
Db 1028 -LATVNVGAG---LLQVQGVVYKANTINLT-----NASAVTTPNPVVVT 1068  
QY 1147 SVADEQGNFTVSNPXYSSYDTSKTSVDITPAGENGIT-----TKVNGVYVVRVIDQ 1197  
Db 1069 GAIDNTGN-----ANNGIVTFTGNTSVTGNVGTNALTATVNVG---AGLLQ 1111  
QY 1198 TKG-----LTPKLTVG-----NNNGKGVIVDSQNGQNTITG-LSNT 1233  
Db 1112 VQGVVYKANTINLTNDASAVTFTNPVVVVTGAIDNTGNANGIV--TPTGNTSVTGDIGNT 1169  
QY 1234 LANYTNDKGSVRTEQGNIIKDEKTRAASIVD-----VLSAGFNLOGCEAVDF-----1283  
Db 1170 NALATVNVGAGITLQAGGSL-----AANNIDFGARSTLEFNGPLDGGGAIPYYPKGA 1222  
QY 1284 -----YSTYDVNFADGNATTAKVTYDDTSKTSKVYDVNVDD 1321  
Db 1223 IANGNNAILNVNTKLTASHLTIGTVAEINIGAGNLTIDASVGD-----VTILNAQN 1275

QY 1322 TTIEVKDKKLGKVTKTTTLTSTGTGANKFALSNQATGDAVKASQDIVAHLNLTLSGDIQT--- 1378  
Db 1276 INFRADSVLVSNL---TGVVNNILL-----AADLVAP-----GADEGTVVF 1316  
QY 1379 ---AKGASQANNSAGYV---DADGNK---VIYDS---TDNKYYQAKNDGTVDKTEVA 1424  
Db 1317 NGVGVNLVGSNVAGTARNIGDGGCNKFNLLIYNATVITDDVNLLEGIONVLIKNADFT 1376  
QY 1425 K-----DKLVAQAQTPDGTLL--AQMNKSVKINKEQVNDANKKQGINE 1464  
Db 1377 SSTAFNAGAIQINDATYITIDANNGLNIPAGNIQFAHADQALVQLQSSGNDRTITLGANI 1436  
QY 1465 DNEFVKGLEKRAASDNKTKNAATVVDLNAV-AQTPPLFAGDTGTT---AKKLGTELTIK 1520  
Db 1437 D-----PNDDEGIVI-----LNSVTAGKKLIAG--GKTFFGAHLK-QTILFKG 1478  
QY 1521 -GOTDNTKLTNNIGVVAGTGTGTVKLAKDLTLNLSYNAGGTKIDDKGVSPVDSGGAKA 1579  
Db 1479 AGDCST-----AGTFTNTTNVLDIT-----GQ-----1501  
QY 1580 NTPVLSANGLDLGGKVISNVGKGTKDTDAANVOQLNEVRNLLGLGNAG---NDNAD-----1632  
Db 1502 -----LELG-----ATTANVVLFNDAVQLTQTGNIIGGFLDFENAKNGMV 1539  
QY 1633 --GNQVNIADIKDPNNGSSSSNRVTIKAGTVLGGKGNNDTEKLA---TGGIQVGVVDXDN 1687  
Db 1540 TLNNNVNAGAVQ--NTGGTNGTLLI---VLGASNLNRVNGIAMLKVAGAGNVIAKGGK 1593  
QY 1688 AN-GDLSNVVVKTKDGSKALLATYNAAGTNYLTNNPAPAIIDRINEPOGIRFEHVNDGN 1746  
Db 1594 VKGEIOGTGNT-----LTLPAHFNLTG-----SINKTGOALKLFPMNGGS 1636  
QY 1747 QEPVVOGRNIDSSASGKHSVAIGFOAKADGEAAVAIGROQTAGNOSIAIGDNAQAT--- 1803  
Db 1637 VSGVW-----GTAANSVGDITTAGATSFASSVNAKGTATL 1671  
QY 1804 -GDOSIA---IGTGNVNVKHSAGIDPSTVKADNSVSGNNNOFTQATQTDTFGVGNNI 1859  
Db 1672 GGTTSFANTFTNTGAVTLAK-----GSITSFAKNVTATSF-VANSA 1711  
QY 1860 TVTESNVALGSNSAISAGTHAGTQAKSDGTAGTITTAGA-----TGTVKGFAGQTAVG 1914  
Db 1712 TINFNSLANSN--ITVG-----GTTTLTGANQVYTYGT-GSFTDTTLN 1754  
QY 1915 AVSVGA---SGAERRIQNVAAAGEVSATSDAVNGSOLYKATOSIANATN 1960  
Db 1755 TTFDGAAKSGNILLIKSGSTLDLSGVSTLAL-----VVTATN 1791  
RESULT 6  
B71704  
cell surface antigen (sca3) RP451 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
C:Accession: B71704  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sierhertz-Ponten, T.; Alsmark  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499  
A:Accession: B71704  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2340 <AND>  
A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14908.1; PID:g3866  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: sca3; RP451

Query Match 5.6%; Score 582; DB 2; Length 2340;  
Best Local Similarity 21.2%; Pred. No. 7.2e-15;  
Matches 499; Conservative 313; Mismatches 879; Indels 658; Gaps 119;

Thu Sep 13 14:18:58 2001

QY 25 SHSTGGSCATQGVSVCTLSFARIAALVAVIGATLSSAY-AQKQDTHXIAIGEQNP 83  
Db 252 AYDLGSSSILTDIGNICSLD-----TINVLLGSATFNSTILKATNINLKH----- 297  
QY 84 RRSCTAKADGRARAIGENANAGCOQATAGSSNKTNGSSSLDKIGTDATGOESIAIGGD 143  
Db 298 -NTSLND-DNIIIVIG--NKGNNKDI--LNFKVHGTWLDNEMIIIPAKPT--HGT 347  
QY 144 VKASGDASIAIGSDHLH---DQGNPKPKG-TLINDLINGHAVLK-----ETRSKDN 195  
Db 348 LNFKNATLNGNINLILKESGCHCKTLNLQGNTKVDNLVAFDSVLDSGTSVINGLIDT 407  
QY 196 D-VKYRTTASGHASTAVGMSYAGHESNAFGPATAKSAYSLAVGLAATAEGOSTIAI 254  
Db 408 DCVTFNNSVNG-GTLIIINAKNTISAKILNA--TKAKTOIANLTMN-HPAGDISDIRI 463  
QY 255 GSDATSSSICALGAGTGAQRAQLQGSF-ALGOGSVVTVSDNNSRPAYTPNTQALDPKFOAT 313  
Db 464 -ADNTIYTIDA-----KNGVNLLNNAKIIIFEGADSMALALINTGVTDRTFTIY 512  
QY 314 NN-----TRAGPLSIGSNSIKR--KIINYGAG---VNKTD 344  
Db 513 NNLNOSGDEYIGVIEAKIKVITIANOSGPYTIQODNTHRLKELIVFEGAGDIIIDDTIF 572  
QY 345 VVVAQLEAVVWAKERRITFGQDDNSTDKVIGLNTLTIKGGAETNALTDNNIGVVKHEAD 404  
Db 573 TKLISNSTGOITFNKTLDLGAGN---LAFKKGHTLVVNGVTGSIITWSENNOGIL--TI 627  
QY 405 NSGLKVKLAKTLNLTNLEVTNTLNAFTTVKVGSSSTTAELSDSLTFTQPN----- 456  
Db 628 NSG-----NITGVIGTNELGLKLVNIGADPVTCSANFASVALTNPSVLLADG 677  
QY 457 ---TGQSOSTKTVGVNGVKFTNNAETTAAGTTRITRDKIGFARGDGVDEQAQPYLDKK 513  
Db 678 VTLTGEVTHNNTKGV--LSLGTGSMITQIGTNSAALEKI----- 716  
QY 514 QLVGSAIITDINGIDAGNKKISNLAGSSANDAVTIEOLKAAPKPTLNAGAGISVTPTEI 573  
Db 717 --NIGAGASINISYAGSVTLTDQTSLETLNNDVVVNSNIITAGNSGK----- 765  
QY 574 SVDAKSNVTAPYNTGVKTELENSDGTSDKFSVKCSGTNNSLVTAEHLASYLNEVNRTA 633  
Db 766 LIETGNGGITG--NIGANGAALQ-----EVVFNGTN-----IGGTA 800  
QY 634 DSALQSTVKEEDDDANALIVA-----KDT---TKNAGAVSILKLKG----- 674  
Db 801 NS--QNFVVAH---SAANVITGLITTCALKYKDTGTIIAHGVLGVGDIDFNKAKGFILGD 855  
QY 675 -----NGLTVATKK---DGVTVFGLSQSGLTIGKSTLNDGLTVKQDNEOIQVG 721  
Db 856 GAMIDGSLCNGGVAGTLDIFIGDGNVTQNICADNANSI--STINIOQ-----DNTKNVTI- 908  
QY 722 ANGIKFTNVNGSNPTGIANTARTRDKIGFAGSDGAVTNKPYLDQDKIQVGNVKITNT 781  
Db 909 ANDIFVNIHFTNGGI-LQLGGNLTTHINIDFANGGGLFEFN-----GN-----NT 952  
QY 782 GINAGGKAITGLSPTLPSIADQSSRNTELGNTTODKDKSMAASINDILNFTG-----FN 834  
Db 953 -YNLNAIIVNGQILNAFTNLKASDDTIGTV-----KIINIGQIGTQPNFT 998  
QY 835 LKNNNPIDFVSTYD-IVDFANGNATATVTHDTANKTSKVYVDVN-----VDDTTI 885  
Db 999 IOVNNKNTLVSVNNSNF--GDANSOLILSAPVDOTIKFINNLNETGGGIITLDSNGN 1056  
QY 886 HLTGTDDNK-KLGVKTTKLNTSANGN-TATN-----FNVNSS---DEBALVNAKD 931  
Db 1057 NLTISGNGIKLGSKNELSLINIKGVTVTNDLQIQNHQILNINNGALFDDQSLTSAK- 1115  
QY 932 TAENLNTLAKEIHTTKGADTALQFTTVKVKVDENNNADANAIVGQ-----KNAN 982  
Db 1116 -IKNIN-----IGTVAGGA-----TYTLDAINDNFDLNTSGMVPKHODSILELKNSSNTN 1164  
QY 983 NOVNTLLKGENGLN-----IKTDKNGTFTFGINTTSGL--KAGKSTLNDGG 1027

Db 1165 DHPIITLSALDPCNQFGIILKITDNTKNLTIDNNGNVAYTLGTANHMLKOLTFASIDNGA 1224  
QY 1028 LSKNPTGSEQIOVGADGVKFAKVNNGVVGAGIDGT-TRITRDEIGFTGTNGSLDKSKP 1086  
Db 1225 IALKVGINVENTLNTKDIELNEVNANVLFNKWTITATGNGINGHVDFOGNAGVI----- 1279  
QY 1087 HLSKDGINAGGKIIITQSGEIAQNSHDAVTKGKIYDLKTELENKI-----SSTAKTAQNS 1142  
Db 1280 NLNDDIEIDGSVSTGNVNGTLNFGSGKVTG-----LINNIVMLQACAGDVSLSA 1330  
QY 1143 LHFEFVADEQG---NNPT-VSNPYSYDTSKT-----SDV 1173  
Db 1331 SGNYSITEIOGNNNLTFEANSHLTDTINKTGGQDLNLVINGGVSYSIGANAAVGD 1390  
QY 1174 ITFAG-----ENGITTKVNGVVRVGID-OTKGLTTPKL-----T 1207  
Db 1391 IINAGSVNFTLKSIGNIVISDGMQVNNVATDISGNANNGTLKLNHTPINITST 1450  
QY 1208 VGNNNGKVIDSONGQNTIT-----GLSNTLANVTNDKGSVRT 1246  
Db 1451 LGNNAIG-TIEVANNDVTITGTLQAGNIHFSNATOATLTLAGAASQVTTNITTAGNNIHT 1509  
QY 1247 TE-----QGN--IIKDEKTRAASIVDVLGAGFNLOGN-----EAVDFSTYDVTNFPADG 1295  
Db 1510 LEVTFDFTGNDGIIIGDAN-NRLKSI-----ELTNGCVTINSPIHYSIITANNAQ 1560  
QY 1296 NA---TTAKVTYDDTSKTSKVYVDVNVDDTTI--EVKDKKILGVKTTTLTSTGTGANKFAL 1350  
Db 1561 NVKLNIEGGITYDLGSKIKSLANVQISEDTTIRGDVYSKVLNIDAGKTINFDGRDNNMP 1620  
QY 1351 SNQATGDALV-----RASDIVAHLNLSGD-IQFAGKASOAN-NSAGYVDAD-GNKVIY 1401  
Db 1621 KNLDPDALIDLPLRSLSLFNFTDIKADNLNFPADDTATANKFDVAVIDAHDNGGIL 1680  
QY 1402 DSTONKY-YQAKNDGTVDKTKEVAKDKLVAQAOTPDGTLAQMNVKS-----VI 1448  
Db 1681 KENDNAWLTQEIKNANII---EIASDKFM-----LLQKNIKAATLIADNANLVLL 1727  
QY 1449 NKEQVNDANKKOGINEDNAFVKGLEKAASDNKTNAAVTVGDLNAVAQTPLTFAQDGTGT 1508  
Db 1728 DNVEVNTNLNRDVIDLDA---NYELKYTGNVTHINGLLTI-----ITY---FDTA 1771  
QY 1509 AKKLGETITIKGGQDNTNKLTDNNIGVVAGTGDGTVKLAKDLTNLS-----VNA 1558  
Db 1772 IQKGGHILVSGSNVDMSDL-DNLI-----IKIKAHSIDITNITSOTKHQIVKLETGA 1822  
QY 1559 GGTKIDDDKGVSEVDSGQAKANTPVLNGL-----DLGGKVISNVG-----KGTKDIDA 1609  
Db 1823 IYTPVPQTKVIIIDASEEQNKFKWADANGLVLLTDTGGR--DDTGGRDDTRGRGNTD-- 1878  
QY 1610 NVQQLNEVENLLGLGN-----AGNDNADGNQVNIADIKK--DPN-----S 1647  
Db 1879 -----NGCRDNCDOGNI-SNNSNEAGSSSDKN-XGITDVVPIDPSPILDYTKNNYVAS 1932  
QY 1648 GSSS---NRTVIKAGTVLGGKGNNDTEKATGGIOGVVDKDG---NANG-----DLSNV 1695  
Db 1933 GIANQLINHVKDFGNTTDAGKLLNDLGFMSPNRVTTETDLRLSNRINVANGLVGVLN 1992  
QY 1696 WK-----TQKD-GSKKALLA---TYNAAAGOTNVLTNPAEADIRINEQGITR 1738  
Db 1993 EVENELFTDAINMNDNPTAKEIGNRLEELSDANTVGLNKTNTLLNKK-----IN 2041  
QY 1739 FPHVNDGNEPVPVQGRNGIDSSASGKHSVAIGFOAKADGEEAAVAGROTQAGNQSIAIGD 1798  
Db 2042 LKELNTNQAIIANGDE--DNIVTGIWGSFVGKIKONSKN-ASGYQSTGGGIIGFDY 2098  
QY 1799 NAQATGDQSTAIQGTGNVAG---KHSAGLGDPTVKADNSYVG-----NNNQFDTA-- 1847  
Db 2099 NI-----DNSIVICAAATMADSKVKKHKNKNDKGRDTKAKSIYSIYGLYNLNNFEVEALG 2154  
QY 1848 -----TQDVFVGNNITVTESNSVALGNSAISAGTHAGT----- 1883

Db 2155 VYGRNKKYKRIITITDQIAIGKFINFYSEYELGYNLIS---HRTTITPMFGNRY 2211  
Qy 1894 QAKSDGTAGTTTATAGTGVKGRAG--QTAVGAVSGASGAERRIQNVAAVESATSTD 1941  
Db 2212 ATFKNNGYKENNTTFQNLISIKKYYDKFTILGLNSV-----THYLSQD 2255  
Qy 1942 AVNGSQL-----YKATOSIANATNDELHRIHONENKANAGISSAMAMASMPQAYIPGRSM 1996  
Db 2256 IIIPELHWFNYOCKNKLPNIDARLD-----GIDEPLTITRFPKPAKITYN-- 2301  
Qy 1997 VTGGIATHN 2005  
Db 2302 LGGGISTKN 2310  
RESULT 7  
T31105  
hypothetical protein 2 - Haemophilus ducreyi  
C:Species: Haemophilus ducreyi  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T31105  
R:Ward, C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.  
J. Bacteriol. 180, 6013-6022, 1998  
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.  
A:Reference number: 220984; MUID:99030326  
A:Accession: T31105  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4919 <NAR>  
A:Cross-references: EMBL:AF057696; NID:g3929021; PID:g3929023; PIDN:AAC79761.1  
C:Genetics:  
A:Gene: lspA2  
Query Match 5.5%; Score 570.5; DB 2; Length 4919;  
Best Local Similarity 19.2%; Pred. No. 5.3e-14;  
Matches 484; Conservative 353; Mismatches 914; Indels 771; Gaps 109;  
Qy 2 NHIVKVPNKATGTFMAVEYAKSHSGGSCATQGVGS-----VCT----- 43  
Db 3 NKRYKLIFS KVKVKNLVPVPAENIKSASGSGSSSKIAEDQEPDLSIACSLPLSSSIH 62  
Qy 44 -----LSFARIAALAVLIGATLGSAYAKKDTKHIAIGEQNPPRS- 86  
Db 63 LGLNHSPLKVPKKNLSVLLSLMPAQVWADSSNAIVDHSHGAKQAVDERD-PKNGK 121  
Qy 87 -----GTAKADGGRAI-----AIGENANAQG-GQAIAIGSSNKTGVNGSSLD 126  
Db 122 EKVVVINIAKPD-EQGISDNHFSKFNIPNSAVFNNSIKEGNSQLVGLLGENKNLGSQAQ 180  
Qy 127 KIGTDATQESIAIGGVKASGDASIAIGSDDLHLLDQHGPNKPKGPTLINDL--INGHA 184  
Db 181 TIFNOVTGDOESKISGGLEVFGEKA-----DLFTI-----NPNGVTLNGVKVTINTDR 227  
Qy 185 VLKEIRSKDNDVK-----YRTTASGHASTAVGAMSYAQGHFSNAGFGRATAKSNYS LAV 240  
Db 228 FVASTSEVVEPHIKOLANVQRKVIIGDKGVATNGLS-----HFD----- 266  
Qy 241 GLAATAEQSTIAIGSDSSSLGAIAGCAGTQAQLOGSIALGOGSVVYQSDNNSRPAYT 300  
Db 267 VVAKNIEQGVKISGDSKPAKLANVTAAGNLIY-----DVNTRDV-N 309  
Qy 301 PNTQALDPKFOATNNKAGPISIGSNS-----IKRKIINVAGVN-----KTDVAVN 346  
Db 310 RNT---NPKKPTDNTNRKDNIAISGESAGSMYGRNIKEFIYTDKAGVNHQGVIPAEDDIN 366  
Qy 347 V-----AQLKAVYKWKAKERITFQGDNDNSTVDKIGLNTLTIKGAETNALT 393  
Db 367 ILTDGNSRLNKVYADYVRVY--GKDIELANNGQIHA--DQQLILNATGHVKVINDGSSVIS 423  
Qy 394 DNNTIGV-----VKEADNSGLVKLAKTLNLTNTVNTTTLNATTTVKVGGSSSITTA 443  
Db 424 NNLGISALNTLENATVYANNLSFRVNTDNLNLSKVSARAAD-----LQSGNLNLDKA 479

Qy 444 ELLSDSLFTQPTNGSQSTSKTVYGVNGVKF-----TNNAETAAICTTTRITRDKIGF 496  
Db 480 SVLAHKLTLNISNDVSLN-NOSKLSANNLIKVKVRDLNLSNELSA-----NNLT 529  
Qy 497 ARDGDVDEKQAPYLDKQKLVGSVAITDNGID-----AGNKKISNLAKSSANDAVT 549  
Db 530 NTSNNITLTK-----NKSFTAGNTLNTVNTNLTNDELAAANNLTNLNTRKNTLND 582  
Qy 550 IEQLKAAKPTLNAGAGISVTPEISVDKSGNVTAPTNIIG-----VKTTELNSDGTSD 603  
Db 583 -SKLSANKLDLN-----VTDNVTLSKS-----TLSAGELTFKKVKNVTLNDSE-- 626  
Qy 604 KPSVKSGSTNNSLVTAEHLASVLYNEVN---RTADSALQSFTVKEEDDDANAITVAKDTT 660  
Db 627 -----LAANNLSLNASHNVTLNKSKLSAOKADIKAVNLTLNDTTELAKNLDINSTTI 680  
Qy 661 KNAGAVSIL-----KLKKGK-----NGLTVATKKDGTVTFGLS 692  
Db 681 TNGTGIAGIFANITTEKLNKEKALILAEQNLNFTVNGSHYENKGDIVSKDKATVTFSKN 740  
Qy 693 QDSGLTICKSTLNNDGLTVKDTNEQIQVGANGIKFTNVNGSNPGTGIANATARITRDKIGF 752  
Db 741 SD-----FTSNGSKLVNAQNLKVNVTISQ-----GDDITLIGNVTLNASGT 785  
Qy 753 AGSDGAVDTNKPYLDQDKLVQGVNK-ITNTGINAGGKAITGLSPTLPSIADQSSRIELG 811  
Db 786 FTNSGNLTIVK-----TLDVGDQIQNTNKG-----NLTVG 815  
Qy 812 NTIQDKKS-----NAASINDILNTGF-----NLKNNNPIDF 844  
Db 816 EDLHKSKTKITNDGKLISIKNLNISSEADFINNLTGLIGIEALKIATKGNFTNKEKAILA 875  
Qy 845 VSTVDIVFANGNATTATVTHDTA---NRTSKVVYDVNVDDTTH-----LGTDDN 893  
Db 876 SNLLDLSVAEGKKTFFNGTIESGKNLNTITGAF-LNVDNATIRSFGLNITSTGNVSN 934  
Qy 894 KKLGVKTKLKNKTSANGNTATNFVNSSDEEDALVNAKDIAENLNLTAKEIHHTKTADTA 953  
Db 935 NGTLISNERLNTS-----AANF-TNESNGTVMSNGL-----LNTIAQGNITNKNLIAS 983  
Qy 954 LQTFVTKVDEN--NNADDANAITVGQK-----NANNOVNTLTKGE-----NGLNIKTD 1001  
Db 984 RQQLNLTAADNITNDSNISNKIAVLHSLGNSLNSKDOVYNL-----GETYAGNNISVRAH 1040  
Qy 1002 ---KNGTVFGINTTSGLKAGKSTLNDGGLSIKNPTGSEQIQVGADGVKFAKVNNGVYGA 1059  
Db 1041 QLKNDVKLWGDITTK-----TREGQASYKLYQASNGHGFNDGSSGSGCDLNIK 1092  
Qy 1060 GIDGTTRITRDEIG--FTGTNGSLDKSKPHLSKDGINAGGK-----KIT--- 1101  
Db 1093 FADLKNKLTIVORIGKIYAGROLTFNKS-----NAGGKSEIINRGITNVKNKLSYDS 1143  
Qy 1102 -----NIOGELIAQNSHDATVGGKIYDLKTELEKNKISTAKTAQNSLHFSVADEQNN 1155  
Db 1144 DVSFENNQOSKV-----DLYT-KIFEAKSDIE-----LTFKT--NGTHPVYL-----N 1184  
Qy 1156 FTVSNPYSSYDTSKTSVDITFAGE---NGITTKVKNKVVRVGIDOTKGLTTP---KLTVG 1209  
Db 1185 PKSNNEKKYRSENTEKPKSIGDLINEALSASAPAEIYAYYSGSSSNVINPVSYLAALG 1244  
Qy 1210 NNGKGIVIDSQNGQNTITGLSNTLAN-----VTDKGSVRTTE 1248  
Db 1245 NAN-----NNSNPHYLNLTALKHILGNQWQDLKKOENIKVLKQWEDPKDKGASKMLD 1298  
Qy 1249 OGNIIKDEKTPAASIVDVLNAGFNLOGNEAVDFVSYDTYVNFADG---NATTAKVYD 1305  
Db 1299 ---LYPNTDKKAKIFAGIIRNGNDTISDVESEDFKKYS--KFQNGEWAKNKDTGDSYD 1353  
Qy 1306 DTSKTSKVYDVNVDDTTIEVKDKKIKLVKVTTLTSTGTGANKFALSNOATGDAVKASDI 1365  
Db 1354 STKASEKYYKVENVDHKE-NIDEHKLNGIKHEITVPGVSFEN--LNNKNMDHQDPKLGEI 1410

QY 1366 VAHL--NTLSGGDIOTAKASQ-----ANNSAGYVDADGNKVIYDSTDN-----KYYQAK 1412  
DB 1411 DKSIISELLAQPVTETKSAARSDPRVNDKEDLDNLYRTRLSYINQNNYLGAKEYFNQ 1470  
QY 1413 NUGTVDKTEVAKDKLVAQAQPDGCTLAQMNVKYNKEQVNDANKKQGINEDNFAVAGL 1472  
DB 1471 LDPEDDKLKGIRK-----TGDNYFEHLQTRLIEKVADVADNHLKHLGH-DIALVKL 1521  
QY 1473 EKAASDNKTKNAAVTVGDLNAVAQPLTTFAGDTGTTAKKLGETLT----- 1517  
DB 1522 IDSAS-----IQAKDLNL-----KVGEALTKEQDKDLKDLVWV 1556  
QY 1518 -----IKGGOTDINKLTDNNIGVYAG-TDGF 1542  
DB 1557 KTEVNGOEVLVPOVYLAKOTIEVERKORGVGCTIRAGIIDVKKVDDVRNTGTIAGVGL 1616  
QY 1543 TVK-----LAKDLTNLSNVNAGCTKIDDKGVFVDSG----- 1575  
DB 1617 EAKNKLKNTGDIILSORLSK-----VGKGLSTGVTVYVDETGATKVRKARIKSEGHIIYE 1672  
QY 1576 -----QAKANTPVLVSLANGLDLG-----GKVISNVGKTKDTDA 1608  
DB 1673 TDKDKNVDTIFASELKGNTGQIKAKDLNLDIYETSYKYKELFGKNGFIRGVRVQTSQ 1732  
QY 1609 ANVOQLNEVNLLGLGNAGNDAGNOV-----NIADIKKDPNSGSSSNR-----TVIKAG 1659  
DB 1733 AKSVGTDASDFHLHLSLEGVDVNOTGSLKANRTTGVVKGDFNTKAGKDLFHRQIDVTSG 1792  
QY 1660 TV-----LGGKGNNDTEKLTGGIOGVVDK-----DGNANGDLSNVWVKTK----- 1701  
DB 1793 TVYSASAGGQSGAGISLTDOGVETVNTKTATAGANADVNFEMKRTRETETSITHRSEF 1852  
QY 1702 -----DGSKKALL-----ATYNAAGQ 1717  
DB 1853 NALSGELYVMCKADIGVDINRDVEVIKTPETEAEOKAEBEAKAEKENEASETAKE 1912  
QY 1718 TNYLTNPAAEIDR-----INEOGIRFFHVNDGNQEPVVGQNRGIDSSA 1761  
DB 1913 TEEAENDVAEKDKTKPKFKKLTDEETAAAFETKGEDEFFAAYKAREEE--DRKKGETLSA 1970  
QY 1762 SKHUS-----VAIGFOAKAAGEAVALIGRQ-----TQAG--NOSTAIG 1797  
DB 1971 EQIESTKARDEKETYVELKVGVAEAEHSAADAISNKAQIIDTQNGKQGGVVALQ 2030  
QY 1798 DNAQ-----ATGQDSIAITGTVNVAAGKHSAGIDPSTVK-----ADNSYSVGNNNQFTDAT 1848  
DB 2031 EASDVNLNATGD-----LAGASAKLKFELSTIEKKSRGASDCRSILGGRLLAAR 2080  
QY 1849 QTDVFGVGNNTVTTESNVALGSNSAISAGTHAGTQAKKSDGTAGTTTAGATGVKGA 1908  
DB 2081 GGDII--TLNNYETTENSLSLKARDNVNNSGV-TEQKDESNOSLKVTAGASSGCCGVMA 2137  
QY 1909 GOTAVGVSVCAGSAERRIQNVAAGEVSATSTDAVNGSQL-YKATQ--SIANATNELDHR 1965  
DB 2138 GCCSAG-VSAGVSGS-----YNESNTESTSHNSLLRCKSLRVEAGKDFNLISSNVVDH- 2191  
QY 1966 THQENKANAGTSSAMAMAPQAYIPGRSMVTGCIATHNG---QGAVAVGLSKLSDNGQ 2022  
DB 2192 LHL-DVKGDTNVVSKQDSYSRKRGVNVSVSAGVGVSTAGGARPNGSVGLGVSAENENSK 2250  
QY 2023 WV 2024  
DB 2251 IV 2252

RESULT 8  
E85822  
probable invasin 23135 [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: E85822  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: E85822  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2660 <STO>  
A:Cross-references: GB:AE005174; NID:q12516151; PIDN:AAG57041.1; GSPDB:GN00145; UMGF:  
A:Experimental source: strain O157:H7, substrate EDL933  
C:Genetics:  
A:Gene: Z3135

Query Match 5.1%, Score 521; DB 2; Length 2660;  
Best Local Similarity 20.8%, Pred. No. 1.9e-12;  
Matches 484; Conservative 277; Mismatches 880; Indels 682; Gaps 103;  
QY 23 AKSHSTGGSCATGQGVSVCTLSFARIAALAVILVIGATLSGSAYAKKDKTHIAIGEONQ 82  
DB 546 ADSHST-----ATLTFIAHDAAGNPVIGLVLS-----TRH----- 575  
QY 83 PRRSCTAKADGDRAIAIG---ENANAQGGQAIAGSSNKTVTNGSSLDKI-GTADATQBSI 138  
DB 576 -----EGVQDITLSDWKDNGDSYQILTTCAMSGTL--TLPOLNGVDAKAPAV 624  
QY 139 AIGGDVKAS-GDASTAIGSDDLHLDDQHGNPKHPKGTLLINDLINGHAVLKEIRSSKDNV 197  
DB 625 VNIISVSSRTHSSIKIDKDRY---LSGNPIE-----VTVELRDENDKPV 666  
QY 198 KYRRITASHASTAVGAMSYAOGHFSN---AFGTRATAKASAYSLAVGLAATA-----E 247  
DB 667 KEOQ-----OLMTAVISIDNVKFGVTTDMKETADGVYKATYATYTKGSLTAKLLQMNNE 722  
QY 248 GOSTIAIGSDATSSSIALGATAGTAAQLOGSIALGQGSVVQTSNNSRPA--YTPNTQA 305  
DB 723 DLHTAGIIDANPQSAKIATLSASNNGLANENAAVTVS-VNVADGSGNPINDHTVTFVA 781  
QY 306 LDPKFOATNTKAGPLSIG-----SNSIKRKIINVCAGVYKNTDA 344  
DB 782 LSGSATSFNNQNTAKTQDVNGLATFDLKSQEDNVVEVTLNGLVQKOTLIVSFGUSSIAQ 841  
QY 345 VNV-----AQLEAVVVKWAK-----ERRITF-----QGDNS----- 370  
DB 842 VDLQSKNEVWADGNDSATMTATVRDAKGNLLNDVKVTENNVNSAAKLSQTEVNSHDGIA 901  
QY 371 -----TDVKIQLDNLTLT--IKGAEETNALDNNIGVYKEADNSCLKVKLAKTLNNLTENV 424  
DB 902 TATLTSKNG-DYTVTASVSSGSAQ-----QVIFIGDQSTAAULTSVSPSGDIITVWT 954  
QY 425 TTLNATTTVKVSSSTTAELLSDSLTFQPNQTSQ--STSKTVYGVNGVYKFTNNAETTA 482  
DB 955 APLHMTATLQDKNGN---PLDKKEITFSPNDVNASRFSISNSKGMWDSNGTALASLTG 1010  
QY 483 AIGTTRITRDKIGFARDGVDKQAPYLDKKQLKY---GSVAITDIONGDAGNKKISNLA 539  
DB 1011 TLAGTHMITARLANSVSDT-OPMTFVADKRAVVVVLQTSKAEIINGVD--ETTITATV 1067  
QY 540 KGSSANDAVTIEOLKAAPKPTLNAGAGISVTPTEISYDAKSGN-----VTAPTNYIGVK 593  
DB 1068 KDPFDN---VVKNLVSVFRT-----SPADTQLSLNARTNENGIAEVLTKGTVLGVHT 1117  
QY 594 TE-LNSDGTSDKFSVK--GSGTNSNLVTAHLASYLYNEVNTADASLOSFTVKEDEDDAN 651  
DB 1118 AEAILLNGNRDTKIVNIAPDASNAQVTLNIPAOQV--VTNNSDSVQLTATVKDPSNHPVA 1175  
QY 652 AITV-----AKOTTKNAGAVSILKLKNGKLTATVATKKDGT----- 686  
DB 1176 GITVNTFTMPOVAAVFLENNG-IAITQANGEAHVILKKGKAGHTVTTATLGNNSDAQ 1234  
QY 687 -VTFGLSQDSGLT-----GKSTLNNNDGL-----TVKDTNEOIQOVGANGIKFTNVN-GSN 734  
DB 1235 PVTFVADKDSAVVVLQTSKAEIINGVDETTLTATVKDPD-----NAVKDLQVTESTN 1288

Qy	735	PG-TGIANTARITRD-----KIGAGSDGAVDYNKPYLDDQLGVGNVKIWTGINAGGKA	789
Dy	1289	PADTQLSQSKSNTNDSGVAEVTFKGTVLGVTAETLPNGN---NDTKIVNIAPDAASNAQ	1345
Qy	790	ITGLSPPLPSIADOSSRNIELGNTODDKSNAASINDILANTGFNLKNNNPDPFVSTYD	849
Dy	1346	VT-LNIPAAQOVVNNSDSVOLATWKDPNSHPVAGIT-----VNFMPDOVAANFT	1395
Qy	850	IYDFANGNATTATVTHDTANKTSKVYDVNVDDTTIHLTGDDNKGLGVTKLNKTSTAN	909
Dy	1396	L----ENNGIAITQANGAEHVTLK-----GKKAGTHTV-----	1424
Qy	910	GNTATPNVNNSDEDALVNAKDIAENLTLAKEHTTKGTADTALOTFTVKKVDENNAD	969
Dy	1425	--TATLSNNNTSQPVTEFAD-----KTSALVLIQISKNEITCNV	1464
Qy	970	DANAITVGOKNA--NNOVNTLTLLKGENGLNIKTDKNGTVTFPGINTSG--LKAGKSTLNDG	1026
Dy	1465	DSATLATATVKDQFDNEVNU-----PVFS--TASSGLTUTPGESNTNES	1507
Qy	1027	GLSIKNTPTGSEQVQGDGVKFAKVNNNGVVGAGIDGVTTRITRDEIGFTGTNGSLDKSKP	1086
Dy	1508	GIA-----QATLAGVAF-----GGQVTVAS-----LANNGASDNKTV	1539
Qy	1087	HLSKDGINAGKKITNIQSGETA---QNSHDVATGCKIYD-----LKTELENKISSAKT	1138
Dy	1540	HFIGDTAAAKIIELTPVPDSIIAGTPQNSSGSVITATVVDDNNGFPVKGVTVN-FTSNAAT	1598
Qy	1139	AQNSLHEFFSVADROGN-NFTVSNPYSYDYTSKTSVDITFACENGITT-----KYMKGVVR	1192
Dy	1599	AEMTNGQOAVTNBQGKATVYTNTRSSIESGARPDVBASLENGSSLSLSTSYNADAOST	1658
Qy	1193	VGIDOTKGL-----TPPKLTGVNNNGKVIDSQNGQNTITGLSNTLANVTNDKGS	1243
Dy	1659	AHLTLLOALEDTVSAGDTTNLIEVKDNGV-----POQEVTLVSPEGVTPSNNA	1712
Qy	1244	VRTTEOGNIIKDEDKTRAASIVDIVLSAGFNLOGGEADVFTSYDTVNFADGNATTAKVT	1303
Dy	1713	IYTT-----NHDGN-----FYASTAXKAGVYQVTTATLEN	1742
Qy	1304	YDOTSKTSKVYDVNVDDTTIEVKDKKLGY-----KTTTLTSTGTGANKFALSNOATGDA	1358
Dy	1743	XDSMQOOT--VTTYPPNYANAEISLAASKOPVIAANNDLTLTATVADTEGNAIAN-----	1794
Qy	1359	LVRASDIIVAHNLTSGLDIGIOTAKGASQANNSAGYVDADGNKVLYDSTDNKYYQAOKNDGTVD	1418
Dy	1795	-----SEVTF--TLPEDR-----ANFTLCDGCKVVTDT-----EGRAKVTLK	1830
Qy	1419	KTREVA-----KOKLVAQAO---TPDCTLQAMNVKSVINKEOVND	1455
Dy	1831	GTRAGAHTVTASMAGCKSEQLVNVFIADTFLTAQVNLNVTEDEFNFIAN-NVGWTRLQATVTD	1889
Qy	1456	ANKQGINEDNAVKGLEKAASDNKTKNAAVTVGDLNVAQOTPLTFAGDGTGTTAKKUGET	1515
Dy	1890	GNGNPLANEAVTFLPADVSASFLLTCQQGS-AITDINGKAEVTL-----SGTKSGTYPVT	1943
Qy	1516	LTIK-GOTDINK-----LTDNNGVVVAGTGDTFKLAKDLNLNASVAGG	1560
Dy	1944	VSVNNYGVSTKOVTLIADAGTAKLASLTSVSYFVVSTEGAT--MTASVTDANGNPVEG	2001
Qy	1561	TKIDDKGVSEVDSSGOAKAN-----TPVLANSGLDLGGKVISNYSGKGTKD	1605
Dy	2002	IKVNFRTSVTLSSTSVETDDRGAEIFLVTSFVGLKTVKSASLADKPTEVISRLLNAKAD	2061
Qy	1606	TDAANVOOLNEVERNLLGLG-----NAGDNDNAGNOV-----	1636
Dy	2062	INSATITSLEIPEEQVMVAQDVAHVAKHNDOFGNPILAESVTFSAEPPEHMTISQNTIVST	2121
Qy	1637	---NIADIKKDPNSGSSNRVTYIKAGTVLGGKGNND---TEKLATGGIQ--VCVDKDG	1687
Dy	2122	DTHCIAEVTWPBERGS---YMWKASLANGSSYEKDLVVIOOKLTLSSASPPLICNGSPTG	2178
Qy	1688	ANGDSLNVWVKYTKQDGSKKALLATYNAAGOTNYLTNNPAEAIDRIHQGIRF-----	1739

## RESULTS

RESULT  
576109

576109 hypothetical protein - *Synechocystis* sp. (strain PCC 6803)

C; Species: *Synechocystis* sp.

A; Variety: PCC 6803

```
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
```

C;Accession: S76109

R.; Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O., K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, K. *Ann. N.Y. Acad. Sci.* 319: 109-136, 1986.

DNA RES. 3, 109-130, 1990

5. Title: sequence analysis of the genome of the uncultured cyanobacterium synecocyst

A:Reference number: S74322: MUID:97061201

A: Accession: S76109  
A: Reference number: S74322, MOID: S7001201

A: Accession: S70109  
A: Status: preliminary

A: Molecule type: DNA

A; MOLECULE TYPE: DNA  
A: Residues: 1-3039 <KAS>

A; Residues: I-3029 <KAN>  
A: Cross-references: EMRL:D63999: GB:AB001339: NID:01001396: PID:BA010087 1: PID:0101

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
A;Cross-References: EMBL:D63999; GB:AB00339; NID:g1001396; PID:BAA10087.1; PID:g1001396

A, NOTE: THE NUCLEOTIDE SEQUENCE WAS SUBMITTED TO THE EMBL DATA LIBRARY, JUNE 1990

Query Match

Query matrix	3.09;	3.09;	3.09;
Best Local Similarity	21.88;	Pred. No.	3e-12;

Matches 521; Conservative 256; Mismatches 901; Indels 710; Gaps 119;

QY 56 VIGATL-----SGSAYA-----QKKDTHIAIGE-----QNQPR 84

Qy 85 RSGTAKADG--DRAIAIGENAN-AQGQQAIAI-----GSSN--KTVNGSSLKIGTDAT 133

Db 687 LDNTAPAA SITLDANITADDIINIAESGQAIPITGTVGGEFNVGDTVTLTVNDKTF TGAV 746

[illegible]

747 CACCT ESTAWDCOT TVDADYETAA - STAEEDACON  
DB

DB /4/ GAGGLE\$INVPGSDLIVDADLI'IAA'---SIAT'DAAGN-----L 782

```
QY      RSSKUNDAIKRTTA-----SGHASTAVGAMS-----ZT06
          | . | | | | | | | | | | | | | | | | | | | |
```

Db 783 GSATDNCQTVTVDTTAPIPTITVNDVTADNITNAAESGOAIPITGTVGGEFENVGDTVTITV 842

QY 259 TSSSLGAIAG-----ACTRAQLQG-SIALGQGVVVTQS-DNNRSP-----AYTPT 303  
Db 903 PIVTINIAVDDIINAVEAGSPVSGITGVEDGVVTVTIDGNTYATVGTGNAWTFNI 962  
QY 304 QALD-PKQADNTNTRAGPLISGSIKRIINVGAGVKNKTDVAVVQAQLEAVVKAERRI 362  
Db 963 PVADIANFEATEEVVATVSDLAGNPATPATRI-----TVDTVAPA-----V 1004  
QY 363 TFGQDDNSDVK-----IGLDNTLTITKGGAETNALTDNNIGVVVKEADNSGLKVLAKTLNN 418  
Db 1005 TIDISDDTGAQANDFITWDDTLVNGFRAEDST-----VVVSLDG----- 1045  
QY 419 LFEVNTTTLNATTTVKVGSST-----TABELSD-----SLATFQPTQSQSTSKTV----- 466  
Db 1046 -IEIGTVTAN-----GAGETLDYGTLLADGDYELSVTATNPTGNSATATQTVVDT 1097  
QY 467 ----YGVGVKFTNNAETTAAGITRITRDKIGFARDGDVDEKQAPYLDKKQKLVGSVAI 522  
Db 1098 TAPTVTINIAVDDIINAVEAGSPVAVSGTTTG-VEDGQV-----VTV 1139  
QY 523 TID-----NGIDAGNKKISLAKGSSANDAVTIEOLK-----AAKPTLNAGAGISVTPT-- 571  
Db 1140 TIDGNTYATVGTGNAWTFNIPVADIANFEATEEVVATVSDLAGNPATPATRIITVDTTAP 1199  
QY 572 -----EISVD-----AKSGNVT-----APTNYGVKTTTELNSDGTSDKFSVK 608  
Db 1200 TVTINIAVDDIINAVEAGSPVAVSGTTTGVEDGVVTVTIDGNTYATVGTGNAWTFNIP 1259  
QY 609 GSGTNSLVTAHLSYLVNEVRTADSALQSFVKEEDDDANAIV-----AKD 658  
Db 1260 VADIANFEATEEVVATVSDLAGNPATPATRIITV-----DTVAPAVTIDISDDTGAQAND 1315  
QY 659 TTKNAGAVSLKLGKNGLTATVKDGTTFGLSQDSGLTIGKSTLNNDLGLVKTNEOI 718  
Db 1316 FITNDDTLF-----NGTAEA-----DSTVVVSL-----DGIEIGTVTANGAGETWLDYGTGL 1363  
QY 719 QVANGIKFTNVNGSPNGIGIANFARITRDKIGFAGSDGAVDTNKYLPDQDKLOGVNVI 778  
Db 1364 L--ADGDYELSVATNP-TG--NSATATQTV-----VDTTAPTVTINIAVDDI-- 1408  
QY 779 TNGTINA--GGKAITGLSPLTSPIDAGQSSRNIEL-GNTIQDKDKSNAASIN-----DILNT 831  
Db 1409 ----INAVEAGSPV-AVSGTTTGVEDGVVTVTIDGNTYATVGTGNAWTFNIPVADIAN- 1462  
QY 832 GFNLKNNNPIDFVSTDIV-----DFANGNATTAT--VTHDTRANKSKVYVGVNVD-- 882  
Db 1463 -----FEATEEVVATVSDLAGNPATPATRIITVDTTAPT-VTINIAVDDIIN 1509  
QY 883 -----TTIHLTGTDNKKKLGKVTTKLNKTSANGNTAT-----NPNVNSD----- 922  
Db 1510 AVEAGSPVAVSGT-----TTGVEDGVVTVTIDGNTYATVGTGNAWTFNIPVADIANFEAT 1565  
QY 923 EDALVNAKIDIAENLTLAKEIHTTKGTADTALQTFVTKVDENNADANAITVGOKNAN 982  
Db 1566 EEVVATVSDLAGNPATPA-----TRNITVDTVAPAVTIDISDDTGAQANDFIT----- 1614  
QY 983 NQVNTLLKGENLNKIDKNGTVTFGINTSGLKAGKSTLNDGGLSIKNPTSEQIQVG 1042  
Db 1615 -NDDTLVFNG-----TAEADSTVVVSLD---GIEIGTVTANGAGETWLDYGT-----LL 1660  
QY 1043 ADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFTGNSGLSKPHLSKOGINA--GGKKI 1100  
Db 1661 ADGDYELSVATNPTGNSATATQTVVDTTAPTVTINAI-----AVDDIINAVEAGSPV 1714  
QY 1101 -----TNIQSGETAQ-----NSHDAVTGKIVDLKTELEN--KISSTAKTAONSLHEFS 1147  
Db 1715 AVSGTTTGVEDGVVTVTIDGNTYATVGTGNAWTFNIPVADIANFEATEEVA-----T 1768  
QY 1148 VADEQGN-----NFTVSN--PSSYDTSKTSVITF--AGE-----NGITTKVNGK-VV 1191  
Db 1769 VSDLAGNPATPATRIITVDTTAPTVTINIAVDDIINAVEAGSPVAVSGTTTGVEDGVV 1828  
QY 1192 RVGIDOTKGLTTPKLTVCNNNGKGI-VIDSONGQNTITGLSNTLIANTYNDKGSVRTTEOG 1250

Db 1829 TVTID--GNTYATVGTGNAWTFNIPVADIANFEAT-----EEVATVSDLAGNPATPATR 1881  
QY 1251 NIIRKDEKTRAASIVDLSAGFNLOGGCEAVDFYSTVDTVNFADGNA---TTAKVTYDDT 1307  
Db 1882 NI--TVDTVAPAVTIDISDDTGAQAN-----DFITNDDTLVF-NGTAEADSTVVVSLDGI 1934  
QY 1308 ---SKTSKVYVNVDDPTTIEVKDKKLGKVTTLTSTCTGANKFALSNOATGDLALVRSAD 1364  
Db 1935 EIGTVTANGAGETWLDYGTLLADGDYELSVTATNPTGNSAT-----ATQTVVDTT 1987  
QY 1365 IVAHLNTLSGD--IQTAKASOANNSAGYVDADGNKVLVYDSTDNKYKQAKNDGTV----- 1417  
Db 1988 PTVTINIAVDDIINAVEAGSPVAVSGTTTGVEDGVVTVTIDGNTYATVGTGNAWTFNI 2047  
QY 1418 -----DKTKEV-----AKDKLVAQAQTPDGTTLAQMNVKSVINKEQVN- 1454  
Db 2048 PVADIANFEATEEVVATVSDLAGNPATPATRIITVDTTAPTVTINIAVDDIINAVEAGS 2107  
QY 1455 -----DANKKQGINEDNAPVKLEKA--ASDNKTNAAVTVGDGL 1491  
Db 2108 PVAVSGTTTGVEDGVVTVTIDGNTYATVGTGNAWTFNIPVADIANFEATEEVVATVSD 2167  
QY 1492 NAVAOTPLTFEAGDTTAKKLGELTITKGGOTDTNK-----LTDNNIGVWAGT--DGFTV 1544  
Db 2168 AGNPATPAT-----RNITVDTVAPAVTIDISDDTGAQANDFITNDDTLVFNCTAEADSTV 2223  
QY 1545 KLAKDLTLNLSYNAGGT--KIDDKGVSVFVSSGQ-----AKA 1579  
Db 2224 VVSLDGI EIGTVTANGAGETWLDYGTLLADGDYELSVTATNPTGNSATATQTVVDTT 2283  
QY 1580 NTPVLISA-----NGLDLGGKV-LSNVGKGTDPDAAVQQLNEVRNLLGLGNAGNDNA 1631  
Db 2284 PTVTINIAVDDIINAVEAGSPVAVSGTTTGVEDGVVTVTIDGNTYATVGTGNAWTFNI 2343  
QY 1632 DGNQV-----NIADIKKDPNSGSSSNRTV-----IKAGT 1660  
Db 2344 PVADIANFEATEEVVATVSDLAGNPATPATRIITVDTTAPTVTINIAVDDIINAVEAGS 2403  
QY 1661 VLGGKNDNTEKATGGIOVG-----VDKGN-----ANGDLSNVW 1696  
Db 2404 PVAVSG-----TTTGVEDGVVTVTIDGNTYATVGTGNAWTFNIPVADIANFEATEEV 2456  
QY 1697 VKTDGDKSKKALLATVNAAGQTNLYTN--PAEAIDRINEQ--GIRFFHVND----- 1744  
Db 2457 VATVSD-----LAGNPATPATRIITVDTVAPAVTIDISDDTGAQANDFITNDDTLVFN 2510  
QY 1745 GNOE---PVVQGRNGID-----SSAGKHSVAIGFQAKADGAEAAVAGIQTOAGN----- 1791  
Db 2511 GTAEADSTVVVSLDGI EIGTVTANGAGETWLDYGTCTLLADGDYELSVTATNPTGNSAT 2570  
QY 1792 QSTAIGDNAQATGDSIAI-----CTGNVAVAGKHSGA-IGDPSTVKAD-NSVS--- 1837  
Db 2571 QTVVDTTAPTVTINIAVDDIINAVEAGSPVAVSGTTTGVEDGVVTVTIDGNTYATV 2630  
QY 1838 VGN-----NQOFTD--ATQTDVFG-----VGNNTVTES----- 1864  
Db 2631 TGNWTFNIPVADIANFEATEEVVATVSDLAGNPATPATRIITVDTTAPTVTINIAVDD 2690  
QY 1865 --NSVALGNSAISAGTHAGTQAKKSDGAGTITTTAG--ATGTVKGFAGCTAVGAVSVGA 1920  
Db 2691 IINAVEAGSPVAVS-GTTTGEV---DGOVTVTIDGNTYATVGTGNAWTFNIPVADIA- 2744  
QY 1921 SGAERRIQNVAAGEVATSTSDAVNGSOLYKATQSI-----ANATNELD 1963  
Db 2745 -----NFEATEEVVATVSD-LAGNPATPATRIITVDTVAPAVNELD 2784

RESULT 10

C48399

ABC-type transport protein ydha.2 - Escherichia coli

C;Species: Escherichia coli

C;Date: 19-Nov-1993 #sequence\_revision 16-Oct-1998 #text\_change 08-Oct-1999



C:Accession: C48399; D64891; H64891  
R: Moszer, I.; Glaser, P.; Danchin, A.  
Biochimie 73, 1361-1374, 1991  
A:Title: Multiple IS insertion sequences near the replication terminus in *Escherichia coli*  
A:Reference number: A48399; MUID: 92190338  
A:Accession: C48399  
A:Molecule type: DNA  
A:Residues: 464-2020 <MOS>  
A:Cross-references: GB:D85081; NID:g3041754  
A:Experimental source: strain K-12  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBI:88089, NCBI:88090)  
R: Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID: 97426617  
A:Accession: D64891  
A:Molecule type: DNA  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-839; 'LDLPLYFQFSVIT' <BLA1>  
A:Cross-references: GB:AE000237; GB:U00096; NID:g1787665; PIDN:AAC74483.1; PID:g1787668;  
A:Experimental source: strain K-12, substrain MG1655  
A:Accession: H64891  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 'M', 915-2020 <BLA2>  
A:Cross-references: GB:AE000237; GB:U00096; NID:g1787665; PIDN:AAC74487.1; PID:g1787672;  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: ydha\_2  
A:Start codon: GTG

Query Match 5.0%; Score 513.5; DB 2; Length 2020;  
Best Local Similarity 21.1%; Pred. No. 2.6e-12;  
Matches 426; Conservative 266; Mismatches 786; Indels 539; Gaps 97;  
Qy 307 DPKQATNNTKAGPLSTGSGSIKKRIINVA--GVNKTDAVNAQLEAVKWA----- 357  
Db 159 DSVFTYENAD-GTISLQDSNGRKATINLQWIDEANNTVALEGSADGATKQYNHNGEL 217  
Qy 358 -----KERRITQGD-----NSTDKIGLDNTLTKGGAETNALTNNIGV-- 399  
Db 218 VITGDNATVNNKGTVDGKDTGTEINGNNGKVIQDGLDVGSGGHGIDITGDSATVDN 277  
Qy 400 -----VKRADNSGLKVKLAKLNLNLTENTTLNATTVKVGSSSTTAELLSLSLTFTQ 454  
Db 278 KGTMTVDPESMGIQIDGKAIVN-NEGESTITNGGTGTQI-NGDDATANNNGKTVYDGK 335  
Qy 455 PWTGQ-----STSKTVY-----GVNGKFTNNAETTAATGTRIT-RDKIGFARDGDV 502  
Db 336 DSTGTEINGNNGKVIQDGLDVGSGGHGIDITGDSATVDNKGTMVTDPEISIGIQVDGD- 394  
Qy 503 DEKQAPYLDKKLVGSAITID-----NGIDA--GNKISLNKGS-----SANDAV 548  
Db 395 ---QAVVNE-----GSAITNGGTGTQINGDADATANNNGKTVTDGKDTGTEIAGNNGK 446  
Qy 549 TIEQLKAAPLNLNAGISVTPTEISVDKSGNVTAPYNTGVKTTTELNSDGTSDKFSVK 608  
Db 447 VIQD--GDLDVSGGGHIDITGDSATVDNKGTMVTDPEISIGIQI-----DG--DQAI 497  
Qy 609 GSGTNNSLVTAELIASYLNEVNTADSAQSFVKEEDDDANAITVAKDTTKNAGAVSI 668  
Db 498 NEG--ESTITNGGTGTQINGDADATANNNGKTVTDGKDTG-----TKIAGNIGI 544  
Qy 669 LKLGKNGLTATKDGTVTGLSDSLTICKSLTLDNG-LTVKDTNE-OIQVGANGIK 726  
Db 545 VNLQD--SLTVTGGAHGVENIG-----DNG-----TVNNKGDIVVSDTSGISGLVINGEAT 593  
Qy 727 FTN---VNGSNPGTGIANTRITDKIGFAGSDGAVDNTKPYLDQDKLVGNKVTNIGI 783  
Db 594 VSNTGDNVNSNEATGFSITT--NSGKVSLSG-----MQVGDFS-TGVDL 635

Qy 784 NAGKAITGLSPTLPSI-----ADQSSRNTELNTIQQDKS--NAASINDIINT 831  
Db 636 NGNNNSVTAAKDLKLVGQKATGINVSGDANTVNI-TGNVLVDKDKTADNAAEYFDFSV 694  
Qy 832 GFNLKNNNPIDFVSTYDIDVFANGCNATATVTHDTAKNTSKVYDVVDDTTHLTGTD 891  
Db 695 GINYGSDNNVTLDGKLTVVSDSEVTSRQSNLFDGSAEKTGLV--VIGDGNVTMMNG-- 750  
Qy 892 DNKKLVGKTTKLNTKSANGTATNFVNSDDEALVNAKDAENLN---TLAKETHHTTKG 948  
Db 751 -----GLELIGEKNALADGSQVTSRTGYSTSVIVWGESVYVINGDTTISGEF--PLG 803  
Qy 949 TA-----DTAL-----QTFVVKVDE--NNNADANAITVQKRNANOVNTLTUKGN- 994  
Db 804 FAGVIRVQDKALLEGSGATLTMDIDSFHHGTRTPELTYADSGA-KIVNKGTVIEIQL 862  
Qy 995 GLNKTDKNGVITFGINTSLGKAGKSTLNDGGLSIKPTGSEQIOVGADGVKFAKVN 1054  
Db 863 GFAPVTGEN-----TTGINSGISLLONG---KDPAPSPIVLLATNG---GSATNA 907  
Qy 1055 GVVGAGIDGTTTRITRDEIGF---TGTNGSLDKSKPHLSKDGINA-GGKKTINIOSGEIA 1109  
Db 908 GTI-----TGKVTQEHSHVFNKYSTGTSNFTFNNDVSIITGLVAQSNSTIINTSGI 961  
Qy 1110 QNSHDVTTGGKIYDLKTELENKISSTAKTQNSLHFEVSVADEQGNFTVSNPYSSYDTSK 1169  
Db 962 LYGRSGVGLAIAD-----STAENQCKITLDSMMVVDANDTTA 998  
Qy 1170 TSDVITFAGENGITTKVNGKV-VRVGIDQTKGLTTPKLTGVNNGNGKIVIDSONQNTIT 1228  
Db 999 MRDIAS-----NSAIDFGTGVGVGTDYSYG-ACKNATAINQLGGVITVINAGAGMA 1050  
Qy 1229 GLSNTLANVTNDKGSVRTTEQNIILKDEDKTRAASIVDVLVSAGFNLOGNGEAVDFVSTYD 1288  
Db 1051 GASNTVIN---QGTINLEKNGV---DDSLAANTLV---GMAYVEHGTAIN-----D 1093  
Qy 1289 TVNFADGNATTAKVYTDTSKTSKVYDV-----NVDDTT----- 1323  
Db 1094 QTVGINIVNGTQAFYNDGTGTI-VNYGTICTGVQCQSGNEYNNTDFTSLIYTGDTIT 1152  
Qy 1324 -----IEVKDKKLG-VKTTTLTSTGTGANKFALSNOATG-----DALV 1360  
Db 1153 RSETVTLNKSAAVYDKLAGNVNNGTISGQDITVSSGELLENTSGGIINLVKLDKGAVI 1212  
Qy 1361 KASDIAHLNLTSGDIQTAGK--ASQANNSAGYVDADCNKVIYDSTDNKYQ-----AKND 1414  
Db 1213 KNAGVMTNNVDVSGGILANNAGETAITMAG---ADSLVNTTGTINKIYQNAVGFNNS 1269  
Qy 1415 GTVDKTEKAKDLVAQAQTPDGTL---AQMNVKSVINKE-----QVNDAN 1457  
Db 1270 GSV--TGRMMSAGGVFNNGT-DGAILMRGAALTGTAVANNEGTVNLGSSSEGNTGMLEVN 1326  
Qy 1458 KQGINEDNAFVKGLEKAA-----SDNKTNAAVTV----- 1488  
Db 1327 NSAFNNGEFTILDNDKNAVINOSGTYLNTGHMINSSSHNGAVNMMGGNGRFINDTGI 1386  
Qy 1489 -----GDLNAVAQTP-----LTFAGDTGTTAKKILGETLTIKGGQDTNKLTD 1530  
Db 1387 DVSAKSLVSVANNAGDQNAFFWNQDNGVINFDHDSASAVK-----VTHSNFATQ 1435  
Qy 1531 NN-IGVAVAGTGTGVVKLAKD--LTNLSNVNAGGTTIDDKGVSVFVSDSSGOAKANTPVLSN 1587  
Db 1436 NDGIMNISGTGAVAMEGDKNAQLVNGGTINLCTAGTTDTGMIGMDANATADA-VIENN 1494  
Qy 1588 G-----LDLGGKVISN-----VGKG-TKDTDAANVQOLNEVRNLGL 1623  
Db 1495 GTINIFANDSFAFSVLTGTGVHVVNNGTVVIADGV7GSLIKOGDSINVEGMN----- 1546  
Qy 1624 CNAQNDN-ADGNQVNIADIKK-----DPNCGSSSNRTVIKAGT-VLGGKGNNDTE 1671  
Db 1547 GNNGNSSEVHGDYTLDPVPKNTVSVTSGSDEAGSGMNNLNGYVVGTVNNGSAGKLKVN 1606



Qy 1065 TRITDEIGTGTNGSLDKSPHLS-KDGINAGKKITNIQSGET-AONSHDVAVTGGKIY 1122  
Db 2656 LRLTVQRL-----LNOAKGVLAGRDLGLDGGELFNGDGRDLSQNSLSVSLGGVLD 2707  
Qy 1123 DLKTELENKISSTAKTAQ-----NSLHEFSVADEQGNFT-----VSNPYSSY 1165  
Db 2708 NOGGALVSEGLTARAALNRGGTFSSAGALALTQSVALNQGRLLSDAGVTLKGASL 2767  
Qy 1166 DTSKTSVITFAGENGITTKV-----NKGVVRV-----GIDQTKGLTPPK 1205  
Db 2768 DNSR-SGVISAKGAVDIRTVGLDLSNRGGISGNAGITLVAARLQNGQGRYSAGKLLDAN 2826  
Qy 1206 LTVGNNGKGVIDSQNGQNTITGLSNTLAN-----VTNDKGSVRTTEQGNII 1253  
Db 2827 LKGLDQROGGVLLV-SETGV-TLDLNGGTLVNRDGGSLIATPGALLRLQLGAVDNGAGGBIS 2884  
Qy 1254 KDEDKTRAASIVDVLASAGFNLOGN-----GEAVDFVSTVDYTNFADGNATTAKVTYDD 1306  
Db 2885 SDRFTLAASLD--NRGRLIGADSLTLRTAQALD-NSLAGVISGAAG-LDIAAARLDN 2940  
Qy 1307 TSK---TSKVYDVNDV---DTTIE--VKDKKLGVKTTTLTSTGTG-----ANKFA 1349  
Db 2941 SAKGTLASRAGIDLRVDGALDNHAEGTVS GARLTILASASLDSNGKGLLSGNAGLSVATGA 3000  
Qy 1350 LSNQATGDAL-----VKASDIVAHNLTLSGDIQTAKASQANNNSAGVYVDAGNKVI--- 1400  
Db 3001 LDNABGGOLISQGVLDVSSADLDNRGGALSGKQSLRLSAANLDNRGGLTSDGELELTAG 3060  
Qy 1401 -YDSTDNKNYQAKND-----CTVDKTEVAKDKLVAQAQTPDGTFLAQNNVKSV 1447  
Db 3061 RVDSADGEISARGDLRTVRLVORQRLIGERGVSLDLRGGDLNQGGLISARGPLSI 3120  
Qy 1448 INKEQVNDANKKQGINEDNAFVKLEKAAS--DNKTKNAAVTGVDLNAVAQT----- 1497  
Db 3121 ---ERLAVLDNRQG--GEIYSQGFELLARRIDNQCQGRISAGKLRDLADALGNAGAGL 3175  
Qy 1498 -----PLTFAGDT-----GTTAKKLE-----TLTKGGQTDNKLFDN 1531  
Db 3176 LSGMQGLTVTGGSLDNSAGGTLSSKDGELATISLGGALDNHGGALVSKGAQRIDAAASLDN 3235  
Qy 1532 NIGVAGTGDGFTVKLA-----KDLTNLSNVNAGTKIDD-----K 1566  
Db 3236 AQGVSGESDVTLSIAGKLDNGQGLVSAQALSFERDITLLN--NAGG-RINGSLLLK 3292  
Qy 1567 GVSFVDSGQAKANTPVLASGLD--LGKVIISNVGKGTGKTDAAANVOQLNEVRNLLGLG 1624  
Db 3293 GASLDSNDGQ-----LISQGRDLAIGLALVN-----AGAARLASGGDLL-LR 3334  
Qy 1625 NAGDNADGNQVN--IADIKKDPNSGSSNRTVTKAGTVL---GGKGNNDTEKL---ATG 1676  
Db 3335 SASVDNRGKGLVSOGLLEISAGSLDNSASGTLASQADMSLRGLGGALRNQDGLIFSQAG 3394  
Qy 1677 GTQGVDRKDGNGAGDLSNVWYKTKQDKGSKALLATYNAAGOTNYLTNPAEAIDRINEQG 1736  
Db 3395 ALEVQAGSLDNRQGLT-----QAQDNRLRTGGALDNOAGRLDSRAGN----- 3437  
Qy 1737 TRFHVNDGNEPVPVQGRNGIDSSASGRHSAVIGFAKADG-----EAAVAIGRQTOA 1789  
Db 3438 ---LDLQSGSLD--NGAGGVNLNSAKGWLKLVTLGLFDNSAGVTQAOISLEIRAGQGVNRQO 3491  
Qy 1790 QNOSTAIGDQAQTD-----QSIATGTGNV-AGK----- 1819  
Db 3492 GHLSALGDDNRIVTADFQNGGGLYASGLLSLDGQRFNLQNGAAAQGGKGVAGRIDFSLA 3551  
Qy 1820 -----HSG-----AIGDPSIVK-----ADNSYSV-GNNQFT 1845  
Db 3552 GALANRFQLESESELHURAAIINSGSLRGLRSGSTRLVAGDLNWAYGVLESANODL 3611  
Qy 1846 DATQTDVFCVGNNTIVTESNVALGSNSAISAG-----THACTQAKKSDGTAGTTTACAT 1901  
Db 3612 DLQLGSLANAGGRILHTGNGTGFGLDSGVIRAGGELITNGLLDIRASEWTSNSSLVQAGRL 3671

Qy 1902 GTVKGFAQTAVGAVSVGASGAERIQNVAAAGEVSATSTDAVNGS---OLYKATOSIANA 1958  
Db 3672 NLDICTFRQTAEGKLLAVQSETGR-----CGDWSNDGILLASNGSLRLLESLGSGYRGNRA 3725  
Qy 1959 TNELDHRTHQENKANAGISSAWAMASMPQAVIPGRSVVTGGIATHNGOGAVAVGLSKLS 2018  
Db 3726 TSLGDFAL-----NAASLDLGNAAASLA-----GGANYTLGAGNLLVNRGRIT 3767  
Qy 2019 DNGQWVFVKINGSADTQGHVGAAGVAG 2044  
Db 3768 AAGDLV-----ASAAALNNYGTLLGGG 3789  
  
RESULT 12  
T31102  
filamentous hemagglutinin 1 - Haemophilus ducreyi  
C:Species: Haemophilus ducreyi  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T31102  
R:Ward, C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.  
J. Bacteriol. 180, 6013-6022, 1998  
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.  
A:Reference number: 220984; MUID:99030326  
A:Accession: T31102  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-4152 <NAR>  
A:Cross-references: EMBL:AF057695; NID:g3929017; PID:g3929018; PIDN:AAC79757.1  
C:Genetics:  
A:Gene: lspA1  
  
Query Match 4.9%; Score 506; DB 2; Length 4152;  
Best Local Similarity 20.2%; Pred. No. 1.3e-11;  
Matches 494; Conservative 342; Mismatches 864; Indels 744; Gaps 126;  
  
Qy 2 NHIYKVFINKATGTPMAVEYAKSHSTGGSCATQGVGS-----VCTLSFARIAAL 52  
Db 3 NKRYKLIFSKYKNCILVPVAENIKSASGSSSSNKIAEDQEEPDLSLACSLS----- 55  
Qy 53 AVLIVIGATLSSGSAQAQKDKTHIAIGEONPRRSCTAKADGDRATAIGENANAQGOAIA 112  
Db 56 -----PLSSS-----IHGLHNH----- 68  
Qy 113 IGSSNKTVNGSSLDKI-----GTDAQEST--AIGGDVKASGDASIAIGSDDLHLLDQ 164  
Db 69 --SPLKVFEGKSLSVLLSLMPATPLLAQNVAAELNGKVYVDSQHS-----STRIYEQ 120  
Qy 165 HGNPKHPKGTDLNDLINGHAVLKEIRSSKDNVDKYRRTTASGHASTAVGAMSYAOGHFSN 224  
Db 121 KTDNRSKDGIVVVEIAN-----PEYDGVSDN-----RFEKNIPNSAV-----FNN 161  
Qy 225 AFGTRATAKSAYSALVLAATAEGOSTIAGSDATSSSIALGALCAGTAAQLOGSIALGO 284  
Db 162 S-----RTESLSQLVG--KLHANIQLQKEAKLIINQ 190  
Qy 285 GSVVTQSDNNSRPAYTPNTQALDPKFOATNNTKAGPLSIGSNSIKRKLIIV--GAGVNTKD 343  
Db 191 ----VTGDHES-----NIQ-----CALEVACKKADLIIVNPNGITLNGVK 226  
Qy 344 AVN-----VAQLEAVVKNKERRITFGDDNDSTDKIGLDNTLTIKGGAETNALTNNIGV 399  
Db 227 TINTDRFVVSTDIIPHRENGILLSVRNG-----KVTID-----KGGVATNGLSHFEV-V 274  
Qy 400 VKEADNSGLKVKLAKTLN--NLTEVNTT-----TLNATTVKVGSSTTAEILLSDLFT 453  
Db 275 ARNIDQKG-KITVAKTENQKSVNPNANITFAAGSLNKLKTREATPISSGTSSTSPAIS 333  
Qy 454 QPNTGSQSTKTVGVN-----GVKFTNNAETTAAGTTRITRDRKIGFAR 498  
Db 334 ADSAGS-----MYGSNKFVVYDKGAGVKHKGIIIFSEN-DINIKMDGNGNASLAKELYAKK 386  
Qy 499 DGDVDEQAPYLDKKQLKQKVGSVAITDNGIDAGNKKISNLAKSGSANDAVTIEQLKAAP 558

Db 387 DILAKDIETKGOLOANNKIIL-----NSTGKINLRNASEVSDNPNVNVKSENALENA 442  
Qy 559 TLAGAGISVTPTEISVDKSGNVTAPTNYIGVKTETELSDGTS---DKFSVKGSGTNN 615  
Db 443 SMSANS--LDVIVTKIEVN--RSSKVSAGTANI--KASNIITLDGSSVANKIIL--NVTNNA 496  
Qy 616 LVTAHLAS----YLN-----EVRNTADSALOSFTVKEE-----DDDDANAIT 654  
Db 497 TLNQSLSAKDMELNVTHNITLNNTSKLSAOKANIKTENLTNLCRASLVAEKLDINAL- 555  
Qy 655 VAKDTTNAGAVSILKLGKNGLT--VANK-----KDGTVTFGLSDSGLTI--GKSTLNND 707  
Db 556 --DKITNNGTIA-----GUTANITTKALENRDNALILA--HQLNFTVNGSHVYVNG 603  
Qy 708 GLTVKDTNEQIQVGANGIKFT-----NVAGSN--POTGIANTARITRDKI 750  
Db 604 DIVSKD--KALVTFNSNDSFTSNGSKLVDAQNLTNVNMFNITOGSEIILHGNVTLNK 661  
Qy 751 GFAGSDGAVDTNK-----PYLDQKLVQV--NWKI--TWGINAGSKAITGLSPTLPSI 800  
Db 662 GNFTNSGNLTMTKELNISIESFINAGNLTGKNLEVHSNNTVKNDOGLVSIENLMSISK 721  
Qy 801 ADOSSRNIEGL-----NTIQDKDKSNAASINDILNTG-----FNLKN 837  
Db 722 TDFTNNGTLGLEALKTASGNGFTNASNGSLASNKSLDIYGNFTNNGTIESVKSLNITN 781  
Qy 838 N-----NNPIDVSYVIDVF--ANGNATV---ATVTHDTANKTSKVYDVYVNDVDTIHLT 888  
Db 782 NYTFINNAT--IKSYGVNLTISOGNFTNDSNGVTMSHDLNITSQ-----ANIKNKLAG 835  
Qy 889 GTDDNKKLGKVTTKLNTKSANGNTA-----TFNVNVSDEEDALVNAKDI---AENLNTLA 940  
Db 836 GOGN-----LTKAGNITNDSNSTALAVLHNSNDINLNNKYNVIGEIYSQAGNISVEA 890  
Qy 941 KEIH-----TTK--GPADTALQFTVVKVDENNADANAITVQKNANNQVNTL 988  
Db 891 KLLHNDVKLSGNTITTKSGNA-----TVKNTSIOGGHLDANSIRVE-----L 934  
Qy 989 TLKG-----ENGLNKTDKNGTIVTFGINT---SGLKAGKST-----LNDGGLSINKPT 1034  
Db 935 TLNGKFAOLDN--QULKVALRKIYAGSNLTTPKAKEGEKEQKSTAGAKIINRGITNVNKNL 992  
Qy 1035 --GS-----EJOIYQV-----ADGVKFAKVNNGVVGAGIDGTTRIT 1068  
Db 993 EYGSNVVDENNRMSQVNLIEKIFGNPNITLTLKNGVTFKDFSNRRRASNDG----- 1047  
Qy 1069 RDEIGFTGNSLDSKPHLSKDGINAGGKITTNIQSGEIAQNSHDVATGKI---YDL- 1124  
Db 1048 -----EGTNKKTFDNVAHLIEEAF-----SGYSNGNDRASDDGHVKSPIYLL 1090  
Qy 1125 -----KTELENKISSATAQNSLH--EFSVADEQGNFTVSNPYSS----- 1165  
Db 1091 VLAQAVNTEGENYL---KALQHIFGPNWMDLTNTNDDTTINDKNQLKWKFKNN 1146  
Qy 1166 -DTSKTSVDITPAGENGITTKVKNVGVVGVGIDQTKGLTTPKLTVGNNGKGVIDS--- 1220  
Db 1147 GENNHSINLIYPADEGVKAKIFAGVLNNG---TNGVEDKVIYQELNDAKAKKEYEDKFAK 1203  
Qy 1221 -----ONGONTITGLSNTLANVTNDK--GSVRTTEQGNIIKDE----- 1256  
Db 1204 KFOGRKPSRQNGEFDWAG--DWAKEGNEYSYKSETEEYKNGIKKEHTVNIKGKHEIKVP 1260  
Qy 1257 -----DKTRASIV-----DVSAGFNLOONGEAVDFVSTY 1287  
Db 1261 TVSFENLNNIHQODKSDGDKSIISSELLAQPIYVAKADVPDVPRAQNDKXAVDEGLY 1320  
Qy 1288 DT-----VNEADGNATAKV-----TYDDTSKTSKVYVYDVNVDDTTI-----EVKDKKLG 1332  
Db 1321 RTRLSYIN--QNYNLGAKYFFNQLDTEDDKLGKIGRIGNYFPHQLITRLIEKVADNHLT 1378  
Qy 1333 VKTTTLTSTGTGANKFAL-----SNQATGDALYK-----ASDIVAHLNT- 1371

Db 1379 LK-----HGLHDIALVKKLIDSASIOAKDLNLKVGELTKEQKDNLKEDIVWYVKT 1430  
Qy 1372 -----LSGDIQTAKGASQANNAGVYDADGNKNVIYDSTDNKYQAKNDG 1415  
Db 1431 VNAOEVLPQVYLAKQVIEBEKORGVGTCOIRAGIIV-----KVDDVRNTG 1478  
Qy 1416 TVD--KTKEVAKDKLVAQAQTPDGTLAQNMVSVINKEQVND-----ANKKQ 1460  
Db 1479 TIAGVAVGLEAKKNL-----KNTGDILSORLSKLKVGKGLLESTGVTVYVDETGATVKRKA 1533  
Qy 1461 GINEDNAFV-----KGLEKASDNKTKNAATVGDNLNA--VAOTPLTF-----AGD 1504  
Db 1534 IKSEGHYLETDKDNVDLTFASELKGNTGQIKAKDLNLDIYETSYKYKELFKNGGE 1593  
Qy 1505 TG-----TTAKKLGCT-----LTIKG--GOTDTNKLTDNNIGVAVGDTGFTVKLAKD 1549  
Db 1594 IGRVTOFSQAKSVGTDAFPDLHLHLSLEGVNDQTSNLSKANRTTGVYVG--DFNTKAGKD 1651  
Qy 1550 L--TNLNSVNAG-----GTKIDDKGV--SFVDSGGOAKANTPVLSANGLDLG 1592  
Db 1652 LFHRQIDVTVTCTVYSASASGGOSAGISLTDQGVETTYTNKTATAGANADV----- 1702  
Qy 1593 GKVISNVCKGKTDDAANVQQLNEVRNLLG-----LGNAGNDNADGNQ--VNI-----ADIKK 1643  
Db 1703 ----TNFKRTRREYETSLTHRNSEFNALSGELYVMGKADIGGVINDRVDVEIKTPEETAA 1758  
Qy 1644 DPNSSSSNRVTVIKAGTVLGGKGNNDTEKLGATGGIOGVVDKDGNGANGDLSNVWVKTKD 1703  
Db 1759 EQKAAEEAKKAEVK-----ENEASETAAKETE--EAENDNVAEKDKTKPKFKKLT- 1807  
Qy 1704 SKALLATVNAAGOTNLTNNPAEAIDR-----INEQIRFEHVNDGNQEPVVGQNRNID 1758  
Db 1808 --EETAAAEFTGCEFFPAAYKAREBEDRKKGFTLSAEQIESTKARDEKETTYYELKVG 1865  
Qy 1759 SSASGKHSVAIGFOKA-----DGEAAVAIGRQTQAGNQSTAIAGDNAQATGQ 1806  
Db 1866 ABAEAHSAADAISNKARQIIDTQGLKQDGTVAL-----QEASDVNLATGLAGSACL 1921  
Qy 1807 STAIGTVNVVACKHSCAIGDPTSVKADNSYSVGNNNNQETDQTOTDFVFGVGNNTVTSNS 1866  
Db 1922 KPELST---IEKSRGA-----SDGRSILGRLNLA--ARGGDI--TLNVTETNSH 1967  
Qy 1867 VALGSNSAISAGTHAGTOAKKSDGTAGTTTGTAGATGTGKGFAGOTAVGAVSAGSABER 1926  
Db 1968 LSLKARDNVNVSGV--TEQKDESNQSOLKVTACASSCGGVAGGCSAG--VSAGVSGS--- 2022  
Qy 1927 IQNVAAGEVSATSDAVNGSQL--YKATQ--STANATNELDHRTHQENKANKAGISSAMAM 1983  
Db 2023 -YNESNTESTSHTNSLIRKSLRVEAGKDFNLISSNVVDVH--LHL--DVKGDNTNVYSKQDS 2079  
Qy 1984 ASMPQAVIPGRSMVTGGIATHNG---QGAVAVGLSKLSLSDNGQWV 2024  
Db 2080 YSRKRGVNVSVSAGVGVSTAGARPNGSVGLGVSAENENSKIV 2123

RESULT 13

T34434  
hypothetical protein K06A9.la - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T34434  
R:Giesel, C.; Gattung, S.  
submitted to the EMBL data Library, December 1996  
A:Description: The sequence of C. elegans cosmid K06A9.  
A:Reference number: 221525  
A:Accession: T34434  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-2232 <E>  
A:Cross-references: EMBL:U08046; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.la  
A:Experimental source: strain Bristol N2; clone K06A9  
C:Genetics:  
A:Gene: CESP:K06A9.la

```

A:Map position: X
A:introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

Query Match      4.8%; Score 494.5; DB 2; Length 2232;
Best Local Similarity 18.0%; Pred No. 1.6e-11;
Matches 385; Conservative 326; Mismatches 958; Indels 467; Gaps 72;

QY 17  MAVA EYAKSHSTGGSCATQOVGSVCTL-----SFARIAALAVLVIGATLGSAYAAQKQD 71
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 252 LPTASSASSSPSAASSTTPVLLSSSTIQSSSGTFPSSVASSPSTVGST-SGAASSSSYA 310
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 72  TKHTAIGEONPRSGTAKADGRARIAIGENANAGQCATAGSSN-----KTV 120
Db    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 311 TVSTIAG-----STGSTITPVPGSSSTIGSTSPSASSSSSGSTWSTI 351
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 NGSSLDKI-----GRDATGQESIAIGDVKASGDASIAIGSDLLHLLDQHGNPKPKGTLI 176
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 352 SGSTGSTVTVPGSSSTFASSTPTASSSSPGSTTVVAPGSS----- 393
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 177 NDLINGHAVLKEIRSSKNDNDVKYRRTTASGHASTAVGMSYAQGHFNAFTRATAKSA 236
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 394 -----TYGSSTPSASSSSS-CTMSINSGSTGSTVTPVPVSSSTF 431
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 237 SLAVGLAATREGOSTIAI--GSDAT--SSSLGATAGAGTRAQLOGSIALQGGSVVTQSD 292
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 432 GSSTPIASSSSSGSTVTVVSGSSSTYGSSTPSSASSAGTASTISGS---TGSTATIVP 487
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 293 NNSRPATVPTOALDPKFAQNTNKAGPLSTIGNSIRKKIINVCAGVKNKTDVAVVAOLEA 352
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 488 GSS--SSVGSSTQASPSPGTMTSGP--TGST-----VTVVPGSSTSPAPS----- 532
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 353 VVKWAKERRITFGDDNSTDVYKIGLDNLTIKGAETNALTDNNIVVKEADNSGLKVKL 412
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 533 -----SSPNPSSSPASTGSTITISGSSSI-----IVSTVSGS----- 564
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 413 AKTLNNTEVNTTLNATTVKVGSSSTTAELLSDSLTFQPNGTGSGSTSKTVYGVNGV 472
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 565 --TVSGTGTSQSTLASSTATPGSSSTVPSSPPSQSPAPNPTGSTITPSQT--SSQSP 620
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 473 KFTNNAETTAAGTTRITROKIGFARDGDVDEKQAPYLDKKQKLVGSVAL-----TIDN 526
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 621 SPSPNPSSPTGSGSTIITPEGSTASPGTSGSTF-----SVATEVTSQSTVPS 671
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 527 GIDAGNKKISNLAKGSSANDAVTTIEQLKAAPTLNAGAGISVTPTEISVDAKSGNVTAPT 586
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 672 GSSLGIGTSSNPSPSSLSPTSCMSTLTSEP-----SPSSQTQSSCAQSTLTTPS 721
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 587 YNIGVKTEILNSDGTSDKFSYKSGGTNSLVTABHLASYLNEVNRITADSAQSFTVKEED 646
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 722 PNPQSSTSSLES---STSGATTSSGAGTMTSPSSSVGS-SQGSTSPAASSTSGEMT 777
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 647 DDANAITVAKDTTKNAGAVSILKCKNGLTVATKDGTVTFGLSQDSGLTICKSTLNN 706
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 778 SQGSTQTPGSSVTSAAILLTSTQOSVTNPGSTVTPRPSTVSGSTSSGSTVTVGSTEAST 837
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 707 DGLTVKDTNEQIQVANGIKFTNVNGSNPGTGIANARITRDKIGCFAGSDGAVTNKPYL 766
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 838 SGSSVASSSPAPTSQN-----PNPST-----SSGSSMITQSPYP 872
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 767 DQDKLQVGNKYITWGINAGGKAITGLSPTLPSTADQSSRNIELGNTIQDKDKSNAASIN 826
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 873 SQSTSPVSSSTPSPG--SPQTTLTSTSPS-----PSQSTTIGST----- 910
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 827 DILNTGFNLKNNNPIDFVSTYDIDVFANGNATVATVTHDTANKTSKVYDVNVDDTTIH 886
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 911 -----QGSTSP--GISTSEEMTSQSTQTPGSTGTVTQPS-TVSDSTSSGSTV 958
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 887 LTGTDDNKLGVKTTKLNKTSANGNTATFNVNNSDDALVNAKDIAENLTAKETHTT 946
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 959 VGSTEGS-----SSP1PSTSQNTNPSTSGSSMSTQTP-----OSSQSTSPVESST 1005
Db    : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 947 KGTADTATQTFVTKVVDENNADDANAITVQKKNANQVNTLT-----LKGENGINI--- 998

```

Db	1006	GATSSSGPCTITLSI--SPSPSPSTIGSSOGSTSPVYSTISOGSTETPGSTGTWKP	1063
Qy	999	-----KTDKNGTVFVGINTTSGLKACKSTLNDGGLSIKNPTGSEQIQVAGDGVKFAK--	1051
Db	1064	STVSGSASSGATMKGSTASSTSGSGSTSPNPSQSTSPSTSGATSPGSGSGTTLTSLSP	1123
Qy	1052	--NNNGVVGAGIDGTTTRIDREIGFTGTNGSLDKKPHLSKDGINAGGKKITNIQSGBIA	1109
Db	1124	SPSQSSTIGSSOGSTSPVSTTSGDMTSQGSTQIP-----GSGTSTQTPSTG--	1171
Qy	1110	QNSHDVATGKIVDL-KTELENKISHTAKTAQNSLHEFSVADEQGNNTTVSNPYSYDTS	1168
Db	1172	--SGSTSEIGTSOGSTQTPRSSLSTSPASTSTQQSVYSTNPSGS--TVQPSTVRGST	1227
Qy	1169	KTSVDITTFAGENGITTKVNGKVVVRGID-----QPKGLTTPKLTAVGNNGKGI	1216
Db	1228	SSGTVTGTGTEGSSSTGGSSATLSSSSPVPSTSQSPNPSTSGSGTP--TPNPSQSTSP	1285
Qy	1217	VIDSQNGQNTITGLSNTLANV-----TNDKGSVRT--TEOGNIILKDEKTRAA	1262
Db	1286	VVSTTTGEMTSHGSTQTPPTIGSTVTPQSTVSGSNSSGTVTIGSSEASTSGSFSKTPS	1345
Qy	1263	SIVDY-----LSAGFNLQNGEAVDFVSTYDVPFA-----DGNATKATVYDDTSKT	1310
Db	1346	SISPVPSTSPITPASSTSGSTISDVSSVSTTSLAPLSSSLPSTVPSPSTQSFSSTSEG	1405
Qy	1311	SKVVYDVNDVTTIEYKDKKLK--VKTUUTLTSTGTCANKFAL-----SNOATG	1356
Db	1406	SSKASSPVPSTQSTPNTPTGTESTSTLLSSTIGSTQHTTMSKASSGSTSPSTNSQTG	1465
Qy	1357	DALVKASDIVAHLNLTSG-----DIOTAKGSAQNNAGYVDADGNKVYIDSTONKYYQA	1411
Db	1466	STVTMGSSSTSGVSTSSASTQPMQSTQSGSAGSTVASSTASPAASSTAFSPSTCT--MS	1524
Qy	1412	KNDGTVDKTEVAKDLVAQAQPDGTQLAMNVKSVINKEQVNDANKKOGINEDNAFVKG	1471
Db	1525	TSSGTGVT--ISESSTTASQSTGSTVTVMGSSST-----SGVSTSSA--SS	1568
Qy	1472	LEKAASDNKTKNAVTVGDLNVAOQPLAFAGDTCTAKKLGELTILKGGOTDNKLTDN	1531
Db	1569	TPQMSTSQSSAGSTVASSTAGLVSTSTVPSTG-----MGST-----	1608
Qy	1532	NIGVAGTDGFTVKLADLTNLNSVWAGETKIDDKGVSPVDSGQAKANTPVLASGLDL	1591
Db	1609	-----SSGTVGSTISESTTASASQSTGTVTMGSSSTSGVSTSSASTQPMQSTSQGSSA	1664
Qy	1592	GGKVISVNGKTKDTDAAVQQLNEVRNLLGLGNAGNDNAGNQVNIADIKKDPNSGSS	1651
Db	1665	GSTVASST-----TGLVSTSTVPSTGTMGSTSSGTG-----STISESTTASAS	1710
Qy	1652	NRTVIKAGTVLGGKGNNDTEKLATGGIQVQVDKGNANGDLNVWVKVTKQDGSKALLAT	1711
Db	1711	SQT---GSTVT--MGSSSTSGVSTSSASGQPMQSTSQG-----	1744
Qy	1712	YNAAGQTNV--LTNNPBAIDRINEQIRFFHVNDGNOEPVQWRNGIDSSAGKHSVAIG	1770
Db	1745	--SSAGSTVWSTASPAAS-----STAFSSGTMSSTSSG	1777
Qy	1771	FQAKADGEAAVAIGRQTOAGNQSIAGDNAQATGDQSTIAIGT--GNVWAGKHSAGID--	1826
Db	1778	TVGSTMQSSTAAATTSHTGS--TVTLGSSSTSSNQMSSTSQGSSVGTVASSTAGLVSTST	1836
Qy	1827	-PS---TVKADNSYSVGNNOFTDATQTDVFGVGNNTVTVTESNSVALGNSAISAGTHAG	1882
Db	1837	VPSSTGTMGSTSSGTGSTTISESSTTASASQSTGTVTMGSSSTSGVSTSSASSTQPMQS	1896
Qy	1883	TOAKKSDGTAGTTTAG--ATGTVKCFAGOTAVGAVSVCASCAERRIQNVAAGEVSATST	1940
Db	1897	TSQSSAGSTVASSTAGLVSTSTVPSTG--TMGSTSSGTGSTTISESSTAASSTSSOTGS	1954
Qy	1941	DAVNGSOLYKATOSIANATNELDHR-IHONENKANAGISAMAMASMPQAV-IPGRSMVT	1998

Db 1955 TVTIGS-----TSQTNPSRSLQITITPSPQSSTESTQTSLSLSPSPSPSTHSVS 2005

Qy 1999 GGATATNGOGAVNG--LSKLSNDGOW--FKINGS 2031

Db 2006 SREGTTMSGATTGDKMSFLSSTGTVTFSSRGSS 2041

RESULT 14

203099

mucin, submaxillary - pig

N:Alternate names: apomucin

C:Species: Sus scrofa domestica (domestic pig)

C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 02-Jun-2000

C:Accession: T03099 #sequence\_revision 24-Mar-1999 #text\_change 02-Jun-2000

R:Eckhardt, A.E.; Timpte, C.S.; DeLuca, A.W.; Hill, R.L.

J. Biol. Chem. 272, 33204-33210, 1997

A:Title: The complete cDNA sequence and structural polymorphism of the polypeptide chain

A:Reference number: Z14839; MUID:98070526

A:Accession: T03099

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-13288 <ECK>

A:Cross-references: EMBL:AF005273; NID:g2581863; PIDN:AAC62527.1; PID:g2581864

R:Eckhardt, A.E.; Timpte, C.S.; Abernethy, J.L.; Zhao, Y.; Hill, R.L.

J. Biol. Chem. 266, 9678-9686, 1991

A:Title: Porcine submaxillary mucin contains a cysteine-rich, carboxyl-terminal domain in

A:Reference number: A40009; MUID:91236743

A:Accession: A40009

A:Molecule type: mRNA

A:Residues: 12139-12167, 'T', 12169-13288 <EC3>

A:Cross-references: GB:M61883; NID:g454837; PIDN:AAA30998.1; PID:g164374

R:Timpte, C.S.; Eckhardt, A.E.; Abernethy, J.L.; Hill, R.L.

J. Biol. Chem. 263, 1081-1088, 1988

A:Title: Porcine submaxillary gland apomucin contains tandemly repeated, identical seque

A:Reference number: A28528; MUID:88087170

A:Accession: A28528

A:Molecule type: mRNA

A:Residues: 12139-12167, 'T', 12169-12641 <TIM>

A:Cross-references: GB:M21174; GB:J03512; NID:g164321; PIDN:AAA30990.1; PID:g552360

A:Experimental source: submaxillary gland

R:Eckhardt, A.E.; Timpte, C.S.; Abernethy, J.L.; Toumadje, A.; Johnson Jr., W.C.; Hill,

J. Biol. Chem. 262, 11339-11344, 1987

A:Title: Structural properties of porcine submaxillary gland apomucin.

A:Reference number: A92606; MUID:87280230

A:Accession: B29789

A:Molecule type: protein

A:Residues: 1572-1607 <EC2>

C:Superfamily: pig submaxillary mucin

C:Keywords: tandem repeat

Query Match 4.8%; Score 491.5; DB 2; Length 13288;

Best Local Similarity 20.0%; Pred. No. 2.2e-10;

Matches 464; Conservative 303; Mismatches 970; Indels 579; Gaps 96;

Qy 8 IFNKATGTMVAEYAKSHSTGGSCATQVGS--VCTLSFARIAALAVLVIGATLGS- 64

Db 1492 IYLOSTGISRTTBE---SGSTLSVIGSGKTSSTILVTSARSSVAGSGTIG-TVSGAS 1547

Qy 65 -----AYAKQKDKTHIAI-GEQNOPR-RSGTAKADGDRAITAGENANAOG- 107

Db 1548 GSTSSSGSPGATRASIGQKTRRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGV 1607

Qy 108 -----GQAIAGSSNKTVNGSSLD-----KIGTDATGQESIA-IGGDVKASGD 149

Db 1608 TETARPSVAGSGTGTVSGASGSGTSSSGSPGATGASIQPETSRIISVAGSSGAPAVSSG 1667

Qy 150 ASIAIGSDLLHLLDGHGNPKRPHKGTILNDLNGHAYLKBEIRSKDNDVKYRRTTAS-GHA 208

Db 1668 ASQAAGTSG-----AGP-----GTTASSV-----GVTETARPS-----VAGSGTGTGTVSGA 1708

Qy 209 STAVGAMSYAQGHFSNAGFTRATKSNYSIAVGLAATAEGOS-----TIAIGSDATSSSUG 264

Db 1709 SGSTGSSSGPGATGASIQPETSRIISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVG 1768

Qy 265 AI-----ALGACTRAQLOG-----SIALGQGVVVTQSDNNRSPATP 301

Db 1769 VTETARPSVAGSGTGTVSGASGSGTSSSGSPGATGASIQPETSRIISVAGSSGAPAVSS 1828

Qy 302 NTOALDPKFOATNNTKAGPLSIGSNSIKRIINVAGVKNKTDVAVVA--QLRAVYKWAKE 359

Db 1829 GAS-----QAAGTSGAGPGTTASS-----VGVETETARPSVAGSGTGTGTVSGASG 1872

Qy 360 RRITFGDDNSTDVKIGLDNT--LTIKGAETNALDNNIGVYKKEADNSGLKVLAKTLN 417

Db 1873 STGSSSGSPGATGASIQPETSRIISVAGSSGAPAVSS--GASQAAGTSGAGPGTTASSV 1929

Qy 418 NLTEVNTTTL--NATTVKVGSSSTTAEILLSLT---ETQNT-----GSQSTSKTV 466

Db 1930 GVTETARPSVAGSGTGTVSGASGSGTSSSGSPGATGASIQPETSRIISVAGSSGAPAVS 1989

Qy 467 YGVNGVAFPTNNA---ETTAAIGTTRITRDKI-GFARDGDVD-----EKQAPYLD 511

Db 1990 SGASQAAGTSGAGPGTTASSVGVVTETARPSVAGSGTGTVSGASGSGTSSSGSPGATGAS 2049

Qy 512 KKOLKVGSAVITIDNGIDAGNKKISNLAKGSSANDAVTIEQL---KAAKPTLNAGAGISV 568

Db 2050 IQPETSRIISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGVVTETARPSV-AGSGTTG 2108

Qy 569 T-----PTEISVDKSGNVTAPTNYIGVKT----- 593

Db 2109 TVSGASGSGTSSSGSPGATGASIQPETSRIISVAGSSG--APAVSSGASQAAGTSGANGP 2165

Qy 594 -TELSNDCGTSD--KFSVKGSGTNNSLVTAEE-HLASYLNEVNRRTADSALQSTVVKKEDDDD 649

Db 2166 GTTASSVGVVTETARPSVAGSGTGTVSGASGSGTSSSGSPGATGASIQPETSRIISVAGS 2225

Qy 650 ANAIVTAKDITKNAG-----AVSILKLGKGLVATKKGDTFTFGLSQDSGLTIG 700

Db 2226 SGAPAVSSGASQAAGTSGAGPGTTASSVGVVTETARPSVAGSGTGTGTVS-GASGSGSSG 2284

Qy 701 KSTLNDGLTGVTKDTEQIQGVANGI-----KFTYVNGSNPGT-----GIANTARTIR 747

Db 2285 SPGATGASIQPETSRIISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGVTEA---R 2341

Qy 748 DKIGFAGSGCAVDYTNKPYLDQDKLVGNVKTNTNGINAGGKAITGLSPTLSIAQDSRN 807

Db 2342 PSVAGSGTGTGTVSGAS-----GSTGSSSGPGATGASIQGP-----ETSRI 2382

Qy 808 IELGNTIQDKDKSNAASINDILNTGFNLKNNNPIDFVSTYDIVDFAN---GNATTATV 863

Db 2383 SVAGSSGAPAVSSGASQAAGTSGAG-----PGTTASSVGVVTETARPSVAGSGTGTGTV 2434

Qy 864 THDTANKTSKVYDVNVDDTTIHLTGTDNKKLGKVTTKLNTSANGTATNTNFVNSDE 923

Db 2435 SGASGSGTGS-----SSGSPGATGASIQGP---ETSRISVAGSSGAPAVSSGASQAAG 2483

Qy 924 DALVNAKDIAENLNTLAKEIHTTKGTADTALOTFTVKVDENNNDADNAITVQGNKANN 983

Db 2484 TSGAGPGTTASSVGVVTETARPSVAGSGTGTVSGASGSGTSSSGSPGATGASIQGP----- 2539

Qy 984 QVNTLTLLKGENGLNITDKNGTVTFGINT---TSSGLKAKSTLNDGLGLKNP----- 1033

Db 2540 ETSRIISVAGSSGA-----PAVSSGASQAAGTSGAGPGTTASSVGVVTETARPSVAGSGT 2592

Qy 1034 -----TGSEQIQVGADGVKFAKNNVNGVVGAGIDGTTTRITRDEIGFTGTNGS--- 1080

Db 2593 TGTVSGASGSGTSSSGSPGATGASIQPETSRIISVAGSSGAPAVSSGASQAAGTSGAGPG 2652

Qy 1081 -----LQSKPHPLSKDGIN---AGKKITNTIQGEIAQNSHDVATGCKIYDLKTELE 1129

Db 2653 TTASSVGVTEARPSVAGSGTGTGTVSGASGSGTSSSGS-----PGATGASIQPETSRI 2706

Qy 1130 NKISSTAKTQNSLHEFVSVADEQGNFNTVSNPYSSYDTSKTS-----DVIITFAGENGIT- 1183

Db 2707 SVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGVVTETARPSVAGSGTGTGTVSGASGSGT 2766











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:39:30 ; Search time 47.77 Seconds  
(without alignments)  
1467.887 Million cell updates/sec

Title: US-09-361-619-7  
Perfect score: 10303  
Sequence: 1 MNHIYKVFNKATGTMAVA.....NGSADTQGHVGAAGGPHF 2047

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	627	6.1	2249	1 190K_RICRI	P15921 rickettsia
2	506	4.9	2003	1 YDBA_ECOLI	P33666 escherichia
3	471	4.6	1577	1 HLYA_PROMI	P16466 proteus mir
4	469.5	4.4	1654	1 OMPB_RICRI	Q53047 r outer mem
5	457.5	4.4	1643	1 OMPB_RICRY	P96989 r outer mem
6	456	4.4	1643	1 OMPB_RICPR	Q53020 r outer mem
7	451.5	4.4	1902	1 P2P_LACIA	P15293 lactococcus
8	448	4.3	3591	1 FHAB_BORPE	P12255 bordetella
9	447.5	4.3	1902	1 P2P_LACPA	Q02470 lactobacill
10	443.5	4.3	1656	1 OMPB_RICJA	O06653 r outer mem
11	441.5	4.3	1902	1 P1P_LACLC	P16271 lactococcus
12	430.5	4.2	1902	1 P3P_LACLC	P15292 lactococcus
13	417	4.0	1608	1 HLYA_SERMA	P15320 serratia ma
14	375	3.6	1325	1 YDEK_ECOLI	P32051 escherichia
15	370.5	3.6	2334	1 WAPA_BACSU	Q07833 bacillus su
16	366.5	3.6	1569	1 YPJA_ECOLI	P52143 escherichia
17	360.5	3.5	1848	1 CBPA_CLOCL	P38058 clostridium
18	355.5	3.5	1300	1 120K_RICRI	P14914 rickettsia
19	351	3.4	1567	1 ICEN_XANCT	P18127 xanthomonas
20	339.5	3.3	1286	1 AIDA_ECOLI	Q03155 escherichia
21	329.5	3.2	1176	1 SLAP_BACSH	P38537 bacillus sp
22	323.5	3.1	1849	1 IGAA_HAEIN	P45386 haemophilus
23	312.5	3.0	1296	1 ASAL_ENTFA	P17953 enterococcu
24	310.5	3.0	1025	1 SLAP_CAUCR	P35828 caulobacter
25	309.5	3.0	918	1 YMJB_CAREL	P34487 caenorhabdi
26	305.5	3.0	1702	1 IGA2_HAEIN	P45384 haemophilus
27	304.5	3.0	1592	1 GTF2_STRDO	P27470 streptococc
28	301.5	2.9	1694	1 IGAA_HAEIN	P44969 haemophilus
29	301.5	2.9	2628	1 HAGA_PORGI	Q51845 porphyromon
30	299	2.9	1829	1 FRPC_NEIMB	Q9JYV5 neisseria m
31	298	2.9	1419	1 ALA1_CANAL	O13368 candida alb
32	294	2.9	1228	1 SLAP_BACST	P35825 bacillus st
33	292	2.8	1183	1 CNA_STAAU	Q33654 staphylococ

34 291.5 2.8 1829 1 FRPC\_NEIMC  
35 291 2.8 1039 1 AG43\_ECOLI  
36 286.5 2.8 1243 1 VG37\_BPK3  
37 286.5 2.8 1541 1 IGAL\_HAEIN  
38 285.5 2.8 933 1 SLAP\_CAMEF  
39 285.5 2.8 1322 1 ICEA\_PANAN  
40 285 2.8 1637 1 MRSP\_STAAU  
41 284.5 2.8 827 1 CSG\_HALVO  
42 283.5 2.8 1258 1 ICEN\_ERWHE  
43 283 2.7 1861 1 APU\_THETU  
44 280 2.7 1462 1 GTFD\_STRMU  
45 278 2.7 1140 1 YM96\_YEAST

#### ALIGNMENTS

RESULT 1  
190K\_RICRI  
ID 190K\_RICRI STANDARD; PRT; 2249 AA.  
AC P15921;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE 190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).  
OS Rickettsia rickettsii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=783;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R;  
RX MEDLINE=90354033; PubMed=2117568;  
RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;  
RT "A protective protein antigen of Rickettsia rickettsii has tandemly  
repeated, near-identical sequences.";  
RL Infect. Immun. 58:2760-2769(1990).  
CC -!- FUNCTION: ELICITS PROTECTIVE IMMUNITY.  
CC -!- PTM: GLYCOSYLATED (POSSIBLE).  
CC -!- DISEASE: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PATHOGENESIS OF  
AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M31227; AAA26380.1; -  
DR PIR; A41477; A41477.  
KW Antigen; Repeat; Signal; S-layer; Glycoprotein.  
FT SIGNAL 1 28  
FT CHAIN 29 2249  
FT DOMAIN 212 1180 13 X APPROXIMATE TANDEM REPEATS.  
FT REPEAT 212 286 A (TYPE I).  
FT REPEAT 287 358 B (TYPE II).  
FT REPEAT 359 430 C (TYPE II).  
FT REPEAT 431 505 D (TYPE I).  
FT REPEAT 506 577 E (TYPE II).  
FT REPEAT 578 652 F (TYPE I).  
FT REPEAT 653 724 G (TYPE II).  
FT REPEAT 725 799 H (TYPE I).  
FT REPEAT 800 874 I (TYPE I).  
FT REPEAT 875 949 J (TYPE I).  
FT REPEAT 950 1021 K (TYPE II).  
FT REPEAT 1022 1093 L (TYPE II).  
FT REPEAT 1094 1165 M (TYPE II).  
FT REPEAT 1166 1180 TYPE I (INCOMPLETE).  
SQ SEQUENCE 2249 AA; 224333 MW; A9D6646C089DF087 CRC64;

Query Match		6.1%;	Score 627;	DB 1;	Length 2249;
Best Local Similarity		22.2%;	Pred. No. 1.6e-15;		
Matches 477;		Conservative 233;	Mismatches 754;	Indels 684;	Gaps 108;
QY	97	IAIGENA-----NAOQGOAIAIGSSNKTGVSSLDKIGTDATQGE-----SIAICGDVKA	146		
DB	44	IATNNAAFSNVGNWNNWNEITAAGVAN-----GTPAGQPONWAFYGGDYTV	92		
QY	147	SGDASTAIGSDDLHLHDQHGNPKPKG-----TLINDLI-NGHAVLKEIRSSKNDVDKY	199		
DB	93	TADAADRI-----IKATINAGT--PVGLNITONTVVGSIITKGNLLPVTLAGKSLTLNG	146		
QY	200	RETTASCH-----ASTAVGAMSVAOGHFSAFGRATRAKASAYSLAVGLAATAEGOSTIAI	254		
DB	147	NNAAAHGFDAPADNYTGLGNIALGGAANAIIIOAAPS-----KITLAGNIDGGIITV	202		
QY	255	GSDA-----TSSSIGAIALGAGRAQLOGSI-----ALGQGSVVYQSDNNRPAVT	300		
DB	203	KTDAALNGTIGNINALATVNVGAGT-ATLGGAVIKATTTKLTNAASVLTLTNANA-----	256		
QY	301	PNTQALDPKFOATNNTKAGPLSIG-----SNSIKRKLIINVAGVYKTDVAVNAQL	350		
DB	257	VLTAID-----NTTGGDNVGVNLINGALSOVTDGIGNINSLATISVGAG-----TATL	305		
QY	351	EAVVWAKERRITFOGDDNSTDVKIGLNDTLTIKGAETNALTDNNIGVYKADNSGLKV	410		
DB	306	GGAVIKATTTKLI-----DAASAVK--FTNPVVVTGAIDNTGNANN--GIVTFTGNSTVTG	357		
QY	411	KLAKTLNLEEVNTTLNATVTVKVGSS-----SSTTAELL--SDSLFTPTOPNCS	459		
DB	358	NVGNNT-----NALATVNVGAGLLOVGGVYKANTINLTNANASAVFTTNPVVVT	405		
QY	460	OSTSKTVGVNG-VKFTNNAAETAAICTRTIKGIFARGDGDVDEKQAPYLDKKOLKYG	518		
DB	406	CAIDNTGNANGLVTFNCTNSTVTDIGNTAL-----ATVNVG	443		
QY	519	SVAITIDNG-IDAGNKKISLAK-----GSSANDAVTTEQLKAAKPTLN	561		
DB	444	AGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLINGALSQVT	503		
QY	562	AGAGISVTPTEISVDKSGNVTAFTYNGVKTTELNSDGTSDKFS-----VKGS-----GT	612		
DB	504	GNIGNTSLATISVGA--GTATLGGAVIKATTTKLTDAASAVKFTNPVVVTGAIDNTGNA	561		
QY	613	NNSLVTAHLASLYNEVNTADSALOSFTVKEEDDDANAITVAKOTTKNAGAVSILKX	672		
DB	562	NGIVTFTGNSTVTDIGNT--NSLATISVAGTATLGGAV-IKATTTKLTNAASVLTLT	618		
QY	673	GKNGLTVAATKDDGT-----VTFGLSODSG-----LTIGKSTLNNDDGLTVKD	713		
DB	619	NANAVLTGAIDNTTGGDNVGVNLINGALSQVTGIDIGNTSLATISVAGTATLGGAVIKA	678		
QY	714	TNEQIQVANGIKFTN---VNGSNPGTGIANTRITRDKIGFAGSDGADVTKNPYLDQDK	770		
DB	679	TTTKITNAVSAVKFTNPVVVTGAIDSTGNANGIVT-----FTGNSTVTG-----	723		
QY	771	LOVGNVKTNTGICINAGGKAITGLSPTLPSIADOSSRIELGNTI---QDKDKSNAASIND	827		
DB	724	-DIGNTNALAT--VNVGAGTAT-----LGGAVIKATTTKLTNAASVLT	763		
QY	828	ILNTGFNLKNNNPIDFVSTYDIVDFANGNATTATVTHDTANKTSKVYDYVNVDDTTIHL	887		
DB	764	LTNANAVL---TGAIDNTTGGDNVGVNLINGALSOVTDGIGNTSLATISVGAG-----	814		
QY	888	TCDDNKKLGKVTTLKLNK-----TSANGN---TATNFVNSDDEDALVNAK-----D	931		
DB	815	TATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAVDNTTGGDNVGVNLINGALSQVTGD	874		
QY	932	IAENLNTLAK-----EHTTKGTADTALQTFVKKVDENNADANAT--	974		
DB	875	IG-NTNSLATISVAGTATLGGAVIKATTTKLTNAASVLTLT-----NANAVLTGAIDN	927		
QY	975	-----TVGOKNANNQVNTLTLKGENGLNIKTDKNKNTVTFGINTTSGLKAGKSLTNDGGLS	1029		

DB	928	TTGGDNVGVNLINGALSQVT--GDIG---NTNSLATISVG-----AGTATL--GGAV	972		
QY	1030	IKNPTSEQIQVACADGVKPAKVNNGVVGAGIDGTTTRITRDEIGFTCTN---GSLDKSKP	1086		
DB	973	IKATT--TKUTDAASAVK--TNPVVVTGAIDNTGNANGIVTFTGNSTVTCNVGTNA	1027		
QY	1087	HLSKDGINAGGKITTNIQSGEIAQNSHDVAVTGKIIYDLKTELENKISSAKTAQNSLHEF	1146		
DB	1028	-LATVNVGAG--LLQVOGVVKANTINLRD-----NASAVFTTNPVVVT	1068		
QY	1147	SVADEQGNNTVSNPYSSYDTSKTSVDITPAGENGIT--TKVKNKVVRVGVIDQ	1197		
DB	1069	GAIDNTGN--ANNGIVTFTGNSTVTCNVGTNATVNVG--AGLLQ	1111		
QY	1198	TKG-----LTPPKLTVG-----NNNGKIVIDSQNGQNTITG--LSNT	1233		
DB	1112	VOGVVKANTINLTNANASAVFTTNPVVVTGAIDNTGNANGIV--TFTGNSTVTDIGNT	1169		
QY	1234	LANYNDKGSVTRTEQGNIIKDEDKTRAASIVD-----VLSAGFNLOGNGEAVDF--	1283		
DB	1170	NALATVNVGAGITLQAGGSL-----AANNIDFCARSTLEFNPLDGGKAIPIYFKGA	1222		
QY	1284	-----VSTYDITVNFADGNATTAKVYDTSKTSKVYVDVNVDD	1321		
DB	1223	IANGNNALNVNTKLLTASHLTIGTVAEINIGAGNLETTIDASVGD-----VTILNAQN	1275		
QY	1322	TTIEVKDKKLGKVTTLTSTGTGANKFALSNOATCDALVKASDIVAHLNLTLSGDIQT---	1378		
DB	1276	INFRARDSVLVLSNL---TGVGVNNILL-----AADLVAP---GADEGTVPV	1316		
QY	1379	---AKGASQANNSAGYV---DADGNK---VIYDS---TDNKYVQAKNDGTVDKTEVA	1424		
DB	1317	NGGVNGLNVGNSVAGTARKNIGGGGKENTLLIYNAVITTDVNLEGIONVLINKNADET	1376		
QY	1425	K-----DKLVAQAOTPDGTL--AQMNVKSVINKEOVNDANKKQGINE	1464		
DB	1377	SSTAFNAGAIQINDATYITIDANNGLNLPAGNIQFAHADALVLQNSNGNDRTITLGANI	1436		
QY	1465	DNAFVKGLEKASDNKTKNAATVVGDLNAV-AQTPLTTPAGDTGTT---AKKLGFTLTIKG	1520		
DB	1437	D-----PDNDEGIVI---LNSVYAGKLLTAG--GKTFEGGAHKL-QTILFKG	1478		
QY	1521	-GQTDITNKLTNNIGVAGTGDGTVYKLAKDLTNLNSVNAGGTKIDDKGVSVFDDSQAKA	1579		
DB	1479	AGDCST-----AGTTENTTNIVLDIT-----GQ---	1501		
QY	1580	NTPVLSANGLDLGGKVISNVGKTKTDAAVQOQLNEVRNLLGLGNAG---NDNAD---	1632		
DB	1502	-----LELG-----ATTANVVLFDNAVOLTQTGNIGGFLDFENAKGMV	1539		
QY	1633	--GNOVNIADIKKDPNSGSSSRRTVIKAGTVLGGKGNNDTEKLA---TGGIOGVVDKGN	1687		
DB	1540	TLNNVNVVAGVQ--NTGCTNNGTLI---VLGASNLNRVNGIAMLVKVGAGNVTTAKGK	1593		
QY	1688	AN-GDLSNVVVKTKDQKSKALLATYNAAGQNTLYLTNNPAAEIDRINEOGIRFFHVDGN	1746		
DB	1594	VKICEIQGTGNT-----LTLPAHFNLTG-----SINKTGGQALKLNPNWGS	1636		
QY	1747	QEPVVOGRNGIDSSAGSKHSVAIGFOAKADGEAAVAIGRQTOAGNQSIAIGDNAQT---	1803		
DB	1637	VSGV-----GTAANSVGDITTAGATSEASSVNAKGTATL	1671		
QY	1804	-GQOSTA---IGTGNVYVAGKHSIGAIDPSTVKADNSYSVGNNNQFTDATQTDVFGVGNNI	1859		
DB	1672	GGTTSFANTFTNTGAVTLAK-----GSITSFAKNVYATSF--VANS	1711		
QY	1860	TVTESVAGLSNATISAGTHAGTQAKKSDGTAGTTTGA-----TGTVKGAGTAVG	1914		
DB	1712	TINFNSLAFNSN--ITGG-----GTTILGANQVYTG--GSFTDTLTN	1754		
QY	1915	AVSVGA--SGAERRIQNVAAAGEVSATSTDAVNGSOLYKATQSTANATN	1960		

Db 1755 TTFDGAAGSGNLIKSGTLDLGVSTLAL-----VVTATN 1791

## RESULT 2

YDRA\_ECOLI  
ID YDRA\_ECOLI STANDARD; PRT: 2003 AA.  
AC P33666; P76087; P76088; P76856; P76857; P76859;  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HYPOHETICAL 205.9 KDA PROTEIN IN PAAY-ACPD INTERGENIC REGION.  
GN YDRA OR B1401/B1405.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN SEQUENCE FROM N.A.  
RP STRAIN=K12;  
RC MEDLINE=97426617; PubMed=9278503;  
RX Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,  
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,  
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 28.0-40.1 min region on the linkage map.";  
RL DNA Res. 3:363-377(1996).  
RN [3]  
RP SEQUENCE OF 464-2003 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=92190338; PubMed=1665988;  
RA Moszer I., Glaser P., Danchin A.;  
RT "Multiple IS insertion sequences near the replication terminus in  
Escherichia coli K-12.";  
RL Biochimie 73:1361-1374(1991).  
CC -1- SIMILARITY: TO S.TYPHIMURUM ORF NEAR CYSG (AC P25928).  
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION: THE GENE CODING FOR  
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT  
BETWEEN AMINO ACIDS 839 AND 840.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000237; AAC74483.1; ALT\_SEQ.  
DR EMBL; AE000237; AAC74487.1; ALT\_SEQ.  
DR EMBL; D90778; BAA15009.1; ALT\_SEQ.  
DR EMBL; D90778; BAA18880.1; ALT\_SEQ.  
DR EMBL; D90779; BAA18881.1; ALT\_SEQ.  
DR EMBL; X62680; -; NOT\_ANNOTATED\_CDS.  
DR EcoGene; EG11307; ydaA.  
DR Hypothetical protein.  
KW CONFLICT 489 489 I -> V (IN REF. 2).  
FT CONFLICT 495 495 I -> V (IN REF. 2).  
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;

Query Match 4.9%; Score 506; DB 1; Length 2003;  
Best Local Similarity 21.0%; Pred. No. 3e-11;  
Matches 423; Conservative 264; Mismatches 775; Indels 554; Gaps 96;

Qy 307 DPKFOATNTRKAGPLSGNSIKRKLIINVA--GVNKTDAVNVAQLEAVVKA----- 357  
Db 159 DSVFTYTENAD-GTISLQDSNGRKATINLWQIDEANNNTVALEGVADGATKWOYHNGEL 217  
Qy 358 -----KERRITFOGDD-----NSTDVKIGLDNTLTIKGGAETNALTDNNIGV-- 399  
Db 218 VITGDATVNNGKTTVDGKSTGTEINGNCKKVIQDGLDVSGGGHGIDITGDSATVDN 277  
Qy 400 -----VKADNSGLKVLAKLTNNLTENVTTLNATTTVKVGSSTSTAEALLSDSLTFQ 454  
Db 278 KGTWTTVPESMGIIQIDGDKAIVN-NEGESTITNGTGTQI-NGDDATANNNGKTTVDGK 335  
Qy 455 PNTGSO---STSKTVY-----GVNGVKPTNNAETAAIGTTRIT-RDKIGFARDGDV 502  
Db 336 DSTGTEINGNCKVIQDGLDVSGGGHGIDITGDSATVDNKGKTTMTVPESIGIQVDGD- 394  
Qy 503 DEKQAPYLDKQKLVGSAITID-----NGIDA---GNKKISNLAKGS-----SANDAV 548  
Db 395 --QAVVNE-----GESAITNGTGTQIINGDDATANNNGKTTVDGKSTGTEIAGNCK 446  
Qy 549 TIEQLKAAKPTLNAGAGISVTPTEISVDAKSGNVTAFTYINIGVKTTELSNDSGTSKFSVK 608  
Db 447 VTQD--GDLDVSGGGHGIDITGDSATVDNKGKTTMTVPESIGIQI-----DC--DQAI 497  
Qy 609 GSGTNNSILVTAHLASYLNEVNRATDSALQSFVKEEDDDANAITVAKDTTKNAGVSI 668  
Db 498 NEG--ESTITNGTGTQIINGDDATANNNGKTTVDGKSTG-----TKIAGNIGI 544  
Qy 669 LKLKGNLTIVATKKDGTVTFGLSQDSGTTIGKSTLNNDG-LTVKDTNE-QIQVGANGTK 726  
Db 545 VNLDG--SLTIVTGAHGVENIG---DNG-----TVNKGDIVSDTSGSIGVINGEGAT 593  
Qy 727 FTN---VNGSNPTGIANTARTIRKIGFAGSDGAVDINKPYLDQDKLOVGNVKTITNGI 783  
Db 594 VSNTGDVNVSEATGFSITT---NSGKVSLAGS-----MQVGDFS-TGVDL 635  
Qy 784 NAGGKAITGLSPTPSI-----ADOSRNIELGNTIODKKS--NAASINDILNT 831  
Db 636 NGNNNSVTLAKDKLVGQKATGINVSGDANTVNI-TGNVLVDKDKTADNAAEYFDPSPV 694  
Qy 832 GFNLKNNNPIDFVSTYDIVDFANGNATTATVTHDTANKTSKVYDVNVDDTTIHLTGTD 891  
Db 695 GINVGSDNNVTLDGKLTWVSDSEVTSGSNLFDGSAEKTSGLV--VIGDGTVMNNG-- 750  
Qy 892 DNKKGKVTTLKNTSANGNTATNFVNSSDEEDALVNAKDIAENLN---TLAKEIHTTKG 948  
Db 751 -----GLELIGEKNALADGSQVTSRTGYSTSVIVVSGESSVYVLNGDTTISGEF--PLG 803  
Qy 949 TA-----DTAL-----QTFVTKVDENNADANAITVGOKNANNQVNTLTAKEN-G 995  
Db 804 FAGVTRVDQKALLETSGSATLTMQDIDSEFHH-----GTRTVEIQNG 846  
Qy 996 LNIKTRKNTVTFGINTTSGLKAGKSTLNDGGLSKNPTGSEQIQGVAGVFAKVNNG 1055  
Db 847 FAFVTGEN-----ITGINSGTISLIQNG--KDPAPSPIVLLATNG---GSATNAG 891  
Qy 1056 VVGAGIDGTTTRTRDEIGF-----TGTNSLDKSKPHLSKDGINA-GGKKITNIQSGEIAQ 1110  
Db 892 TI-----TGKTEQHSFVNKYSTGTSNFIENNDVSSITGLVAOSNSTIINTDSCIIL 945  
Qy 1111 NSHDAVTGKGIYDLAKTELENKISSATAQNSLHEFSVADEOGNFTVSNPYSSTVDSKT 1170  
Db 946 YGRSGVGLATAD-----STAENQKTKLDSMWVDANDTTAM 982  
Qy 1171 SDVITFAGENGITTKVNKGV-VRVIGIDQTKGLTTPKLTGVGNNGKGVISDSONGNTIIG 1229  
Db 983 RDIAS-----NSAIDFGTGVGVGTSVSG-AGKNATAINQLGGVITIYNAGAMAAYG 1034  
Qy 1230 LSNLTANTVNDKGSVRTTEQGNIIKDEKTRAASIVDVLISAGFNLOGNGEAVDFVSTYDT 1289

[illegible]



Db 350 QGADLTIDGKOLQKQETDIDNRHFPYSWKYDVTKKEKQIQOISQIDAKNNATLTATKGDV 409  
QY 911 -----NTATFNVNSSDE---DALVNAKDIAENLTLAKEIHTTK---GTADTALQTF 957  
Db 410 TLDAAKINAGNLAIAINANKDIHINGLVEKESRSENGN---KRHTSRLESWSNSHOTE 466  
QY 958 TVRKVD---ENNADANAITV-GOKNANNQVNTLTKGENGUNI-KTKNGTIVTVEIN 1011  
Db 467 TLKASELTAGKDLGDAQSGITAGAKLANENVLVNAKDNILNVQKTTNNDKVTIDNHV 526  
QY 1012 TTSGLKAGKSTLNDGGLSINKPTGSEIQVQAGDS-VKFAKVNNGVVGAGIDGT---TR 1066  
Db 527 MWGIGGQGNKNNQOVSHAT-----QUTDAQQLLAADNNVNITGSOVKGNQGAFAVK 581  
QY 1067 ITRDEIGF-----TGINGSLDKSKPH-----LSKD 1091  
Db 582 TTQGDVVIDNALSETISKIDERTGTAFNITKSSHKNETKQSTGSELISDAQLTVVSGN 641  
QY 1092 GINAGKKITNIOS-----GEIAQNSHDVATGKGIYDLKTELENKISSAKTAQNSLHE 1145  
Db 642 DVNVIGSLIKSADKLGTHSLGIDNVKSAQOVT---KIDDEKTSL---AITGHAK----- 689  
QY 1146 FSVADEQ---GNFT-VSNPYSSVTSKTSVITFAGENGITTKVNGKVVVRVGIDQTKGL 1201  
Db 690 -EVEDKQYSGAFHFTHTNKTSTETEQANS--TISGAN-VDLOANKDVTFAGSDL----- 741  
QY 1202 TTPKLTGVNNN---GKGIVI-----DSQNGQNTITG-----LSN 1232  
Db 742 ---KTTAGNASITGDNVAFYSTENKQTDNTDTISGFSYTGVDKVGSKADFOYDKQH 798  
QY 1233 TLANVNDKGSVRTTEOGNIIKEDKTRAASIVDVLSAGFN-----LQNGEAVDPVST 1286  
Db 799 TQTEVTKNRGS-QTEVAGDLTITANK-----DLLHEGASHIHVEGRYQESGNIQHLAV 850  
QY 1287 YDTVNPADGNATTAKVTYDFTSTSK----VVYDVNVDDTTIEVKDKKL---GVKTTTLT 1339  
Db 851 NDS-----ETSKTDSLNVGIDVGNLDYSGVTKPVKKAIEDGVNTT--- 891  
QY 1340 STGTGANKFALSQAOTGDALVKASDIVAHLNTLS-----GDIDTAKGASQ 1384  
Db 892 ---KPGNNTDLTKKVT-----ARDAIANLANLSNLETPNPGVEVGIKGGSQSOQSOTDSQ 942  
QY 1385 A---NSAGVVDAGNKVIVYDSTDNKYQAKNDGTVDKTEKAKDKLVAQAQPDGTL-- 1439  
Db 943 AVTSINAGRIDIDSNNKLHD--QGTHYQSTQEGI-----SLTANTHTSEATLDK 990  
QY 1440 -----AQMNVKSVINKQVNDANKKOGINEDNAFVKGLEKAASDNKTKNAAVTV 1488  
Db 991 HQTFTHETKGGQIGVSTKTKGSD--ITVAIKRGQOTTDNALMETKAGSQFTSNGDISINV 1049  
QY 1489 GD-----LNVAQTPLTTFAGDT-----GTTAKKLGETILTIK--- 1519  
Db 1050 GENAHYEGAQDAQKQKTVINAGDGLTLAQATDTHSESQSQNVNGSANLKVTGTPESKDYG 1109  
QY 1520 ----GGQDTNKK-LTDNNIGVAGTGDFTVKLAKDLT-----NLNSVNAAGTK 1562  
Db 1110 GGFNAGTTHHSKEQTTAKVQITGSGQIELNAGHNLTLQGHLSSEQDIALNATN---K 1165  
QY 1563 IDDKGVSVFSSQAQKANTPVLSANGLDLGKVISNVGKTKDPTDAANYOOLNEVRNLG 1622  
Db 1166 VDLQAS---SEHTEKGN-----NLSSGVOAGFGKRWMTD-DASSVNGLSAQAPAG 1212  
QY 1623 -----LGNAGNDNADGN-----QVNTIADIKKDPNSG-----SSSNRTVIKA 1658  
Db 1213 KODEKSVSREGGTINNSGLNTINGNSVHLOGAQNKSDTQLTQSQGDIEITSQAQSTYDN 1272  
QY 1659 --GTVLG--GKGNNDT-----PKLAT-----GGIQGVGD-----KGNANGDLSNVWV 1697  
Db 1273 NWGTDIGFNGKTKNTPTKEVTEKPAATSIHIGGKLLVNVVEDQOKTSHQNTLETGTITI 1332  
QY 1698 KTOKD-----GSKKALLATYNAACQTNLTNNPAEADIRNEQIGIR--FFHVNDGNGEPV 1750

Db 1333 NSNKDLTSLGANVTADSVTGNVGGSLNIASOKES---DRHVTVGNNVGYNTNTPKSSQV 1389  
QY 1751 VQGRNGIDSSASG-----KHSVAIGFOAKADGEAAVAIGROTAQAGNSIA-----IGDN 1799  
Db 1390 NK-----TAKAGGSLEKTIKDTIDSGIKSSTD-----AISDKYNSLSSTIADTKTGISDE 1439  
QY 1800 AQATGDOSIA-IGTG--NVVAGKHSIGAIDPSTVKANSDYSVGNNOFTDATQYDVFVG 1856  
Db 1440 TKAKIDQGFQKVGNGIKNIIVTG-----AEGHTANADIKVTHVDNDVATKTT----- 1485  
QY 1857 NNITVTSNSVALGNSAISAGHTAGTQAKSDGCTAGTTTTAGATGTVKGFAGOTAVGAV 1916  
Db 1486 ---SLTSNNDLSNVN-----GSTKLTGA---EIVSQQGQVDLGG 1520  
QY 1917 SVGASGAERRIONVAAGEVSATSDAVNGSOLYKATQSIAN-----ATNE 1961  
Db 1521 SV-----KLENTIEGHYHAEAGADLDLKSSVVVDLAKQLVGGDISFKSPVKINE 1566

## RESULT 4

OMP\_RICRI  
ID OMPB\_RICRI STANDARD; PRT; 1654 AA.  
AC Q53047;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
OS OMPB.  
OS Rickettsia rickettsii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=783;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=R;  
RX MEDLINE=92167802; PubMed=1724278;  
RA Gilmore R.D. Jr., Cieplak W. Jr., Policastro P.F., Hackstadt T.;  
RT "The 120 kilodalton outer membrane protein (romp B) of Rickettsia  
RT rickettsii is encoded by an unusually long open reading frame:  
RT evidence for protein processing from a large precursor.";  
RL Mol. Microbiol. 5:2361-2370(1991).  
RN [2]  
RP SEQUENCE OF 279-1654 FROM N.A.  
RC STRAIN=R;  
RX MEDLINE=90136087; PubMed=2515418;  
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;  
RT "Cloning, expression and sequence analysis of the gene encoding the  
RT 120 kD surface-exposed protein of Rickettsia rickettsii.";  
RL Mol. Microbiol. 3:1579-1586(1989).  
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
CC LAYER WITH HEXAGONAL SYMMETRY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X16353; CAA34403.1; -;  
DR Antigen; S-layer; Cell wall.  
KW CHAIN 1 1333 120 KDA SURFACE-EXPOSED PROTEIN.  
FT CHAIN 1334 1654 32 KDA BETA PEPTIDE.  
FT DOMAIN 1181 1188 POLY-THR.  
FT SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;  
SQ

Query Match	4.6%	Score 469.5	DB 1:	Length 1654																
Best Local Similarity	22.1%	Pred. No. 4.8e-10																		
Matches 386	Conservative 174	Mismatches 561	Indels 629	Gaps 90																
650	ANAITVAKD	TTKNAGAVSILK	KGKNGLVATK	KDGVTFVGLSQDSGLTGIGKSTLN	DGL 709															
19	ASTATIVAS	FAGSAGMAA	IOONRTTG--	AATVVDGA--	GFDQTAAPA-----NVGV 66															
710	TVKDTNEQ	IOVGA-NGIK	FTNVGNSPG	CIANTAR-----	ITRD-KIGFAGSDGAVDTN 762															
67	AL----	NAVITANANG	INFNTWPG	SNGFL--LLTANN	LAVTVSEDTLGF-----TN 115															
763	KPYLDQDK	QVGNVKI	TNTGINTG	AGGK--AITGLS	PTLPSTADOSSRNI	EIGNTIOODKDN 821														
116	-----	VVHNAHSE	NLTLNAGK	TLTITGQGV	TNAQAA--ATKNAO--	NVVVOFNNGA 162														
822	AASINDIL	NTG-----	PNLKNNNP	IDFVSTY	VIDVFANGNAT	TATVTH-----D 866														
163	AIDNNDLK	GVGRIDF	GAPASTLV	FNLANPTQ	KAPLILG	NAVIANGVNGLT	NVNGFTQ 222													
867	TANKTSK	VVYOVN	VDD-----	TTIH	TGTDNDKK	LGVKTTKLN 904														
223	VSNKSFAT	VKAINIAD	GQIIFNTD	NANNANT	LNLAQGG	TTFINFTDGTGRL-----VLLS 278														
905	KTSANGNT	ATNPNVNS	DEEDALV	NAKAIOA	ENMLNTL	AKEIHTKGTAD	TALOTFTVKKVDE 964													
279	KHAA----	ATNFTITG	SUGG--	NLKG	VIE--FNTVAV	DGQLTA-----	313													
965	NNNADDAN	ATVGNK	NNANNV	NTLTLK	GENGLNI	KTDKNCTV	TGINTVGLKAGKSTLN 1024													
314	--NACAAN	AV-IGTN	NGAGRA	GFVVS	DNG--KVATID-----	GQVIYAK 353														
1025	DGSLIKNP	TSEQIO	VADGVK	FAKVN	NNNVG	VGCAGIDG	TRITRDEIGFTG	TNGSLDKS 1084												
354	DMVIO	SANATG-----	QVNFRI	VDVGD	AGTT-----	380														
1085	KPHLSKD	GINAGG	KKTIQ	ISGEIA	QNSHD	AVTGKGI	YDLKTELENKIS	TAKTAQNSLH 1144												
381	-----	AFKTA	AKVT-----	ITQDS	NFGNTD	FGNLAAQ	IKVPNAITL	T-----	418											
1145	EFSVADE	QGNFT--	VSNPYSS	DFSKT	SDVTF-----	AGENGIT	TKVN 1187													
419	-----	GNTG	ASNP-----	GNTAG	VITPD	ANGTLES	ASADAN	VAVTNNITAIEA 463												
1188	KGVRV	RGIDQ	TGKLT	PKTL	TVGN--	--KGVID	SQNCQNTIT	GLSNTLAN	VNDKG 1242											
464	SG----	AGVVL	SGTHAA	ELRL	GNAGS	IFKLAD	GTVING	VKNQ	TALVGG	ALAAAGTITLDGS 520										
1243	SVRTTE	QGN-----	IIKDE	KTRA	SIVD	VSAG--	--FN	LQNGE	AVDF	VSTYD 1288										
521	ATITD	GIGNAG	GAALQ	RLIT	LAND	AKKTL	TLGGANI	IAGG	GTID	LQANGCTIKL	TSTQN 580									
1289	--TVNF	ADGNAT	KAVY	DDT	SKTSK	VYDV	VNDVTT	TEV	KDKL	GVKTT	TLTSTGTGAN 1346									
581	NIWVDF	LA	IA	TQ	GT	GV	PD	ASLT-----	NAQ	TLTG	IKTIGT	IGANKTL	GFENIGSS 633							
1347	KFALSN	-QATG	DALV	KASD	IVAH	LN	FLSGD	IQ	TAKGA--	--SO	ANN	SAGI	VYD	ADGN	KNKIYD 1402					
634	KTVLS	NGN	VAIN	EL	VIGD-----	--GAV	QF	AH	DT	YL	IT	TR	T	NA	AGO-----	GKIIFN 679				
1403	STDN	KYYQ	AKN	DG	TV	DKTE	VAKDK	LVA	QAQ	IT	PD	GT	LQ	M	V	KN	VKS	VINKEQ	VND	ANKOGI 1462
680	-----	PVYN	NGTT-----	LAAG	T	ILG	SAT	NP--	--LAE	IN	FGS-----	--AGV 711								
1463	NEDNAE																			

QY	1553	LSNV	-----NAG-----GTKID-----DKGVSFVDSG--	1575
		:		
Db	832	VNALKQITVSGPCNVVINEIGNAGNHHGAVTDITAFENSSLGAVVFLPRGIPENDGNMT	891	
QY	1576	--QAKANTPVLSGLDILGGKGVNKGKTDTAANVQOLNEVRNLLGLGNAGNDNA--	1631	
		:		
Db	892	PLTIKSTVGKTKAFDGFPSVVVLGV-----DSVIAGQVIGDNIVGLG-LGSDNGII	945	
QY	1632	-----DGNVNIADIKKDPNMSGSSSRNTVIKAGTVLGGK-----	1665	
		:		
Db	946	VNATTLYAGISTLNNNGVTVLGGVPPNTPG-----TVVGLGTGTGASKEFKQVTFTTDYNN	1001	
QY	1666	-GN-----NDTEKATGGTQGVVDKDG-----NANGD-----LSNVVWKTKQD	1702	
		:		
Db	1002	LGNIIATINATINDGVTVTTGGI-AGIGDFDKITLGSVNGNVRFAFDGILSN--STSMI	1057	
QY	1703	GSKKALLATYNAAGQTNILTNPAEAIIDRINEOGIRTFHHVNDGNQE--PVVQGR-NGIDSS	1760	
		:		
Db	1058	GTHKA-----NNGTVTYLGN-----AFVGNIGSDTPVASVRFVTFGSDSG	1096	
QY	1761	AS-----GKHSVAI-----GFOAKAD-----	1776	
		:		
Db	1097	AGLCNIYSOVIDFGTYNIGIVNSNIILGGGTTAINGKIDLVTLWTLFASGTSTMGNNTS	1156	
QY	1777	-----GEAAVAIGRO---TOAGNSQIAIGDNAQA--TGDQSIAGTGNVYVAGK	1819	
		:		
Db	1157	IEWTPLTANGNIHIVILGGAQVNTTGTITIKVODNANANFSGTQTYTLIQ---GAR	1213	
QY	1820	HSGAIGDP-----STVKADN--SYSVGNNOFTDATQTVF-----GVGN	1957	
		:		
Db	1214	FNGTLGPNFAVGTGSNRFVNSLIRANQDYYVITRTNNAENVVNDIANSPFGGAPGVQ	1273	
QY	1858	NIT--VTESNSV-----LGSNSAISA-----	1881	
		:		
Db	1274	NYTFEVNATNTAATNNLLLAKNSANSANFPVGAIVTDTSAATNVQLDLAKDIOAQLGNRL	1333	
QY	1882	GTQAKKSDGTAGTTTTAGATGTGVKGAGQTAIVGAVSVSGASGAERRIQNVAG-EVSATST	1940	
		:		
Db	1334	G--ALRYLGTPEETAEMAG-----PEAGAISAAVAGDEALDNVAVGIWAKFPYT	1380	
QY	1941	DA-----VNGSOLYKA-----TQSIAI-----ATNELDHRIHONENKANA-G	1976	
		:		
Db	1381	DAHQSKKGLGAGYKAKTGVVIGLDTLANDMLMGAIGITFKDIKHQDYKKGKTDVNG	1440	
QY	1977	ISSAMAMASMPQAYIPGRSMVTGGIATHNGQGAVALVGLSKLSDNCQWVFKINGSADTQGH	2036	
		:		
Db	1441	PFSESLYGA-----QQLVKNFPA---QGSIAIFSLNQVKNKSQRYE-----FDANGN	1482	
QY	2037	VGAAVGAGFH	2046	
		:		
Db	1483	MSKQIAAGHY	1492	
RESULT 5				
OMP_RICHTY	OMP_RICHTY	STANDARD;	PRT;	1645 AA.
AC	P96989;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	OUTER-MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)			
DE	(SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (ROMP B)			
DE	(CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)			
DE	(120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).			
GN	SLP.			
OS	Rickettsia typhi.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			
OC	Rickettsiaceae; Rickettsiae; Rickettsia.			
OX	NCBI_TaxID=785;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=WILMINGTON;			
RX	MEDLINE=94040787; Pubmed=9224886;			



Db 1012 -TLTTGG-----IAGTD-FDGKI-----TLGSING-----NANVKEVDR 1043  
 QY 1574 S-----GQAKNTPVLSANGLDCKVLSNNGKGTDFKDDAANVOQLNEVRNLLGLGN 1625  
 Db 1044 TFSHPTSMIVSTKAGQIVTY-----LGNALVGNL-----GSDIPVASVR-----F 1085  
 QY 1626 AGNDNADGNVNIADIKKDPNSGSSNRVIVKAGTVLGGKGNKTEKATGGIQGVGDKD 1685  
 Db 1086 TGNDSVGLGNHNSONID---FGTYNLTILNSDVLGG-----GTTAINGEIDL 1133  
 QY 1686 GN-----ANGDLSNVVVKQDKSKALLATYNAAQOTNLTNNPAEADRINEQIRFEH 1741  
 Db 1134 TNNLIPANG--TSTW-----GNNTSLSTLN-----1157  
 QY 1742 VNDGNEPVVQGRNGIDSSASGKSHSVATGFOAKADCEAAVAIGROTOAGNOSIAGDNAQ 1801  
 Db 1158 VSGNVVQIV-----IAEGAQVNA-----TTTGTITIKIQDNAN 1191  
 QY 1802 A--TGQDSTAIGTGNVAVAGKHGATGDPSTVKADNSYSVGNNGQFT-----DATQTDVF 1853  
 Db 1192 ANFSGTQTVTLIQG---GARFNGTLGAP-----NFDVTGNIFVXYELIRDAQDYVL 1241  
 QY 1854 GVCNNITVTESNVALGNSAI--SAGTHAGTQAKKSGCTAGTTTAGATGTVKGFAGQT 1911  
 Db 1242 TRNDVNLNVVITAV-----GNSAINAPGVHONTAI-----CLESTDAAVNNML--LAKDS 1291  
 QY 1912 AVGAVSYGASGABERRIONVAAGEVSATSDAVNGSOLYKATOSIANATNELDHRITHONEN 1971  
 Db 1292 SDVATEIGA-----IATDGAATVATVNLMDTO--KTQDLGNRLGAL--RYLSNSE 1338  
 QY 1972 KANAGISSAMAMSPAQ-----YIPGRSMVTGGIATINGOGA--VAVGLSKL 2017  
 Db 1339 TADVGGSETGAVSGGDDAIDQVSYGVWAKPFYINAEQDKKGLAGYKARTAGVVGGLDPL 1398  
 QY 2018 SDNCQWVFKINGSADTQGHVGAAGV 2042  
 Db 1399 ANDNLM-----IGAAIG 1410

RESULT 6  
 OMPB\_RICPR STANDARD; PRT: 1643 AA.  
 ID OMPB\_RICPR 30-MAY-2000 (Rel. 39, Created)  
 AC Q53020; Q9ZCM0;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
 DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
 GN OMPB OR SPAP OR SPA OR RP704.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsia; Rickettsia.  
 OX NCBI\_TaxID=782;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=BREINL;  
 RX MEDLINE=91045972; PubMed=2122457;  
 RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;  
 RT "Characterization of the gene encoding the protective paracrystalline-  
 RT surface-layer protein of Rickettsia prowazekii: presence of a  
 RT truncated identical homolog in Rickettsia typhi";  
 RL proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BREINL;  
 RA Moron C.G., Yu X.J., Walker D.H.;  
 RT "Sequence analysis of ompB of Rickettsia prowazekii";  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MADRID E;

MedLINE=99039499; PubMed=9823893;  
 Anderson S.G.E., Zomorodipour A., Andersson J.O.,  
 Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria";  
 RL Nature 396:133-140(1998).  
 RN [4]  
 RP PARTIAL SEQUENCE.  
 RC STRAIN=BREINL;  
 RX MEDLINE=92114896; PubMed=1370573;  
 RA Ching W.M., Carl M., Dasch G.A.;  
 RT "Mapping of monoclonal antibody binding sites on CNBr fragments of  
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia  
 RT prowazekii";  
 RL Mol. Immunol. 29:95-105(1992).  
 RN [5]  
 RP IDENTIFICATION OF CLEAVAGE SITE.  
 RC MEDLINE=92104668; PubMed=1729180;  
 RA Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;  
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer  
 RT membrane protein of rickettsiae: identification of an avirulent mutant  
 RT deficient in processing";  
 RL Infect. Immun. 60:159-165(1992).  
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION  
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; M37647; AAA26390.1; ALT\_INIT.  
 CC EMBL; AF161079; AAD42234.1; -  
 CC EMBL; AJ235273; CAAL5140.1; -  
 CC Antigen: S-layer; Cell wall.  
 CC CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.  
 CC FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.  
 CC FT VARIANT 257 257 V -> A (IN STRAIN BREINL).  
 CC FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).  
 CC FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).  
 CC FT CONFLICT 178 179 AA -> VC (IN REF. 1).  
 CC FT CONFLICT 191 201 TQEAPLTGA -> INSRSSYLHVS (IN REF. 1).  
 CC FT CONFLICT 212 212 T -> I (IN REF. 1).  
 CC FT CONFLICT 313 313 Q -> L (IN REF. 1).  
 CC FT CONFLICT 1104 1104 D -> G (IN REF. 2).  
 CC FT CONFLICT 1123 1123 T -> S (IN REF. 2).  
 CC FT CONFLICT 1123 1123 T -> S (IN REF. 2).  
 CC SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;  
 SQ

Query Match 4.4%; Score 456; DB 1; Length 1643;  
 Best Local Similarity 21.1%; Pred. No. 1.5e-09;  
 Matches 399; Conservative 254; Mismatches 642; Indels 600; Gaps 101;  
 QY 107 GGAIAIGSS---NKTVN--GSSLDKIGTDTATGQESIAIGGVKASGDASIAIGSDDLHL 161  
 Db 27 GFGSVAMGAAMOYNRTTNAATFDGIFDQ-----AAGANIPVAPNSVITANAN--- 76  
 QY 162 LDQHGNPKHPKGTLLINDLINGHVAFLKEIRSSKDNVDKYRTRTAGSHASTAVGAMSVAGQH 221  
 Db 77 -----NP-----ITFNPNGHLSNLSFLDTANDLAVTINEDTTLGTITNIA-----QQAQ 120  
 QY 222 FSNAGFTRATKSAISYSLAVGLATATAGQSTIATGSDATSSLSGAIALGACTRAOQSGIA 281  
 Db 121 FFN-----FTVAAGKIILNITGQGITVQESNTINQALTKVIG--GAINANDL 168





```

QY 1266 DVLSAGFNLOAGBAVDFVSTYDFVNFADGNATTAQVYDDTSKTSKVYDVNVDDTTIE 1325
DB 1300 DMLK-----NSVTFQGVTEGANEF---NATSAKF-YDPKGTIATITGVKHKPTTTLQ 1348
QY 1326 VKDKKLGVK-----TTVLTSTGTGANKFAL-----SNOATGDALVKASDIVAHLNLTSG 1374
DB 1349 VDGKQIPIKDDLTFTSFTLDLGLTGQKPGVGVGDTTQNTKTFQALTFILDAVA--PILSL 1406
QY 1375 DIQTAKGASQANNSAGVYDADGNKVIYDSDN-KYYQAKNDGT-----VDKTRVAKDK 1427
DB 1407 DSST-----DAPVTNNPNEFQITATDNAGYLSLSINGSSVASOYVDININSCKPG 1458
QY 1428 LVAQAQPDGTIAOMNVKSVINKEQVNDANKKQGINEDNAPFVKLEKAASDNKTKNAVT 1487
DB 1459 HMAIDQ-PVKLLGKKNLTV-----AVTD-----SEDNTTKNITVYVPEKTL-AAPT 1505
QY 1488 VGLDNAVAQPLTFAGDTGTAKKLGFTLTIKGGQDTNKLTDNNIGVAGTGDGFTVKLA 1547
DB 1506 V-----TP-----STPEAKTVTLTANSAA-----GETVOYS 1533
QY 1548 KDLTNLNSVNAAGFKIDD---KGVSFYDSSGQAKANTPVLSANGLDLGGKVISNVGKGT 1604
DB 1534 AD-----GGKTYQDVPAAGVT-VTANGTFKFKSTDLYGNESPAVDYVVTNI-----K 1579
QY 1605 DTDAAVVOQL-NEVRNLLGLGNAGNDNAGNQVNIADIKKDPNCGSSSNRTVIRKAGTVLG 1663
DB 1580 ADDPAQLQAQKQELTNL-----IASAKTLSASGKYDDATT-----TALA 1618
QY 1664 GKGNNDTEKLTATGIG--VGVKDKNGNDLSNVVWKTORDGSKKALLATYNAAGOTNYL 1721
DB 1619 AA-----TOKAQALDQTNASVDSLTGANRDLQT-----AINQLAALKLPADKKYSL 1665
QY 1722 TNNPAEAIIDRINEQIRFFHFVNDGNQEPVVGVRNGIDSSASGKHSVAIGFOAKADGEAAV 1781
DB 1666 --NQLQSVKAALETDL-----GNQ-----TDSSTGK-----TFTAALDDLVA- 1700
QY 1782 AIGRQTAGNQSIAIGNAQAQDQSTAIAGTGNVNVAKHSGAIGDPTVVKADNSYSVGN 1841
DB 1701 ---QAQAGTQT---DBQLQAT---LAKVLDAVLAKLAEG-----IKAAATPAEVGNA 1742
QY 1842 NOFTDATOTDVEGVGNNTVTESNSVALGSN-----SAISAGTHAGTQAOKSD 1889
DB 1743 K---DAATGKTWYADIADTLTSGQASDASDKLAHLQALQSLKTKYKAAVPAKTVGKGD 1799
QY 1890 GTAGTTTGTAGATGVKGFAGTAVGAVSVGASGAERRIQNVAAEV-----SATSTD 1941
DB 1800 GTTCTSDKGGGQGTPTPPG-----DIGKDKGDEGSPSSGCGNIPTPNPTTTTSTD 1851

RESULT 8
FHAB_BORPE STANDARD; PRT; 3591 AA.
ID FHAB_BORPE
AC P1225;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE FILAMENTOUS HEMAGGLUTININ.
GN FHAB.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90355839; PubMed=2388559;
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Genetic characterization of Bordetella pertussis filamentous
RT haemagglutinin: a protein processed from an unusually large
RT precursor."
RL Mol. Microbiol. 4:787-800(1990).
RN [2]
RP SEQUENCE OF 1-3261 FROM N.A.

```

```

RX MEDLINE=89202384; PubMed=2539596;
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Filamentous hemagglutinin of Bordetella pertussis: nucleotide
RT sequence and crucial role in adherence."
RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
CC -I- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND
CC INFECTION.
CC -I- SUBCELLULAR LOCATION: SURFACE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60351; AAA22974.1; -
DR EMBL; M60351; AAA22975.1; ALT_INIT.
DR EMBL; M60351; AAA22976.1; ALT_INIT.
KW Antigen; Hemagglutinin.
SQ SEQUENCE 3591 AA; 367420 MW; EF7418B30D6E5138 CRC64;

Query Match 4.3%; Score 448; DB 1; Length 3591;
Best Local Similarity 20.5%; Pred. No. 7e-09;
Matches 486; Conservative 291; Mismatches 948; Indels 640; Gaps 104;

QY 9 FNKATGTFMAVEAKSHSTGG---GSCATGQGVSVCTLSFARIAALAVLIGATLSGS 64
DB 264 YDHATRRATPPIAAGARGAAGAAAYAIIDTGAAGAMTKHHTL-VSSDGLGVQOLGSLSPS 322
QY 65 AYAAKCKDTKHITAEQONQPRRSCTAKADGDAIRAIGENANAGGQAIAIGSNKTVNGSS 124
DB 323 AITVSSQGE-IALGDATVQR--GPLSLKAGVWSAGKLASGGAVNVAGGAVKIASASS 379
QY 125 LDKTGTATGQESTAI---GGDVKASGDASTAIG-SDDLHLLDQHGHPKHPKGTLLINDLI 180
DB 380 VGNLAVGGGKRVQATLLNAGGTLVSGRQAVQLGAASSRQALS-----V 423
QY 181 NGHAVLKEIRSKNDVKYRRTTASGHASTAVGAMSYAQGHFSNAFGTRATAKSAYSLAV 240
DB 424 NAGALKAKDLSAT-----RRVDVQKQAVALGSAS-----SNALSVRAG-ALKA 468
QY 241 G-LAAT---AEGSTIAGSDATSSSLGAIAGAGTGAQIAGGSLTQGGSVVTTQSDNNS 295
DB 469 GKLSATGRDLVDGQAVTLGVSAD---GALSVSAGGNLRANELVSSAQLEVRGQRE--- 522
QY 296 RPATPTQALDPKFEQATNNTKAGPLSTGSIKRIKIIIVGA-GVNTKTDVNVVAOLEAVV 354
DB 523 -----VALDDASSARGMTVVAAGALAARNLQSK-----GAIGVQGGVAVSVANASDA 570
QY 355 KW-----AKERRITFGDDNSTDVKIGLDNTLTIKGGAETNALT 393
DB 571 ELVRGRQVDLHDLASAARGADISGEGRVNTGRARSDSDVKVSAHGALISMTALGAIG 630
QY 394 DNNIGVKEADNSGLKVLAKTLNNLTTEVNTTTLNATTVKVGSSSTT-----ABEL 446
DB 631 VQAGSVSAKD---MRSRGAVTVSGGAVNLGVDQSGQVRATSGAMTVRDVAAADLA 687
QY 447 SDSLTFTPTQNTGSGSTSKTVYGVNGVKFTNNAETTAALGTTRTRD---KIG-FARDGDV 502
DB 688 LQAGDALQAGFLKSGAMTVNGRDAVRL---DGAHAGQLRVSSDGOALGSLAAKGEL 743
QY 503 DEKQAPYLDKKQLKGVSAITIDNGIDAGNKKISNLAGSSANDAVTIEOLKAAKPTLNA 562
DB 744 TVSARAATVAELK-----SLDNISVTGGERVSVQSVNSASRVAI-----SAHGALDV 791
QY 563 GAGISVTPTEISVDKASGNVTAPTNYIG-----VKTTELSADGTSDFKSVKSGSTNNSL 616
DB 792 G-----KVSAKSG-----IGLEGNAGVADSLGSDGA---ISVSG-----RDA 826
QY 617 VTAEHLASYLNEVNTAD-----SALQSFVTKKEEDDDANAITVAKDTTKNAGAVSI 668

```



Db 827 VRVQARS-LADISLGAEGGATLGAVEAGSIDVRGGSTVAANSLSHNRD----- 875  
Qy 669 LKLGKNGLTATKDGTVTFGLSODSGLTGKSTLNNDGLTVKDTNEQIQVGANGIKFT 728  
Db 876 VRVSGKDAYRVTAATSGG---GLHVSSGQDLGAVQARGALALD-----GGAGVALQ 925  
Qy 729 NVNGSNPGTGIANTRITRDKIGFAGSDGAVDNKPYLDQDKLQVGNKYIYNTGAG-- 786  
Db 926 SAKAS---GTLHVQGEHLDTLAAVAGVDN-----GTGDRVAKLVSDAG 972  
Qy 787 ---GKAIT-GLSTPLPSIADQSRNIELGNTIQDKDKSNAASINDILLATGFNLKNNMPI 842  
Db 973 LQAGRSWTLGIVDTTDLQARQOKLEGLSVKSDGGLQAAA----- 1013  
Qy 843 DFVSTYDIVDFANGNATATVTHDTANKTSKVYVDVNDVDDTHILGTDDNKKLGKVTTK 902  
Db 1014 -----CGALSIAAA-EVAGALELSQGVTVDRASARARIDSTGSGVIGALK 1059  
Qy 903 LNKTSANGNTATNFNVNSDEDEALVNAKDIAEMNLTAKEIHTKTADTALQTFVTKV 962  
Db 1060 A-----GAVEAASPRRARRALRODFFTPGVSVVYRAOGNTVVRG 1098  
Qy 963 DENNADANAIVGOKNANNVNTLTLKGENGLNKTOKNGTIVTFGINNTSGL---KAG 1019  
Db 1099 DPHGVLAQGLIDMAKG---GTLRLR-----NDALTENGTVT--ISADSAVLEHSTI 1146  
Qy 1020 KSTLNDGGLSIKPTGSEQIOVGADGVKFAK-----VNNNGVVGAG---IDGT 1064  
Db 1147 ESKISQSVLAAGKDKGPV-----SVKVAKLFLNGLTAVNDNNETMSGROIDVVDGR 1201  
Qy 1065 TRITRDEIGFTGNSLDKSKPHLSKDGINAGGKKTINIOSGEIAQNSHDVAVTGKIYDL 1124  
Db 1202 PQT-----DAVTGEARKDESVDALVADGGPIV-VEAGELVHAGGIGNG----- 1248  
Qy 1125 KTELENKTSIATAONSLEH-FSVADEQG-----NNFTV-SNPYSYDTSKTSOV 1173  
Db 1249 -RNKENGASVTRVTGNLVNKGYSAGKQGVLEGGVGGALTNEFLVSGDGTQRIEQAOTENR 1307  
Qy 1174 ITF-----AGENG-ITTKVNGVVRVGDIDOTKGLTTPKLVGNNNGKGIVIDSQNGONTI 1227  
Db 1308 GTFQSQAPAGTAGALVVKAAEAIHVDGVMAIKG---EMQIAGKGGSGPTVAGAKATISA 1364  
Qy 1228 TGLSLNLANVTN-----DKGSVRTT-----EOGNIKDEDKTRAASIVDVLSAGENLQ 1275  
Db 1365 NKLSVDVASWDNAGSLDIKKGAAQVTVAGRYAEHGEVSIQDGYTVA---DAIALAAQVT 1421  
Qy 1276 GNGEAVDFVSYDT-----VNFADGNATTAKVTYDDTSKTSKVYVDVNDVDT 1322  
Db 1422 QRCGAANLTSRHDTRFSNKKIRLMGPQVAGGPVSNTGNLKVREGVTVTAASFD---NET 1478  
Qy 1323 TTEVDKKLGVKTTTLTSTGTGAN--KFALSNOATGDALVKASDITVAHLNLTLSGDIOTAK 1380  
Db 1479 GRAVM-----AKSATLITSGAARNAGKMQVKEAAT---IVAAS--VSNPGTFTA----- 1522  
Qy 1381 GASQANNSAGYVDAGD-----NKVIYDSTD-----NKYYQAKNDGTVDKTKEV-----AK 1425  
Db 1523 GKDIITVTSRGGFNEGKMEKNKDIKVIKTEQFSNGRVLDAKHDHDLTVTASGQADNRGSLKAG 1582  
Qy 1426 DKIVAQAQTPD--GTL-----AQMNVKSVINKEQVNDANKKQGINEDNAFVKGLEKASD 1478  
Db 1583 HDFTVQAQRINDSGTWAACHDATLAKAPHLRNTGQVVAGHDHIIINSKLENTGRVDARND 1642  
Qy 1479 NKTNAAVT--VGDNLNAVAQTPPLTFAGDT-----GT---TAKKLGTELTI 1518  
Db 1643 IALDVADFTNTGSLYAEHDTLTLAOGTQTDRLVDPQDHLVPAEGTLRVKAKSL--TTEI 1700  
Qy 1519 KGGQ-----TDTNKLTDNNIGVVAGTD-----GFTVTKLAKDLT 1551  
Db 1701 ETGNPGLIAEVENIDNKQAIIVGKDLTSSAHGNVANEANALLWAAGELTVK-AQNT 1759  
Qy 1552 NLNS--VYAGGTTKIDKGVSVFVDSQGAQANTPVLISANGLDLGGKVISNVGKGTQTDAA 1609

Db 1760 NKRAALIEAGGNARLTAAVALLNKLGRIRA-----GEMHLDAPRIENTAKLSGEVQRK 1813  
Qy 1610 NVQQL--NEVRNLLGLG-----NAGNDNADGNQVNIADIKKDPNNGS--SSNRRTVIKAG 1659  
Db 1814 GYQDVGGEGHGRSGGLGYVNYWLRAGNKKAG---TIA---APWYGGDLTAQSLIEVG 1866  
Qy 1660 TVL---GKGNNDTEKLTATGGTQGVQVQDKNANGDLSN----- 1694  
Db 1867 KDLYLNAGARKDEHRLHLLNEGVIQAG--GHGHTGGVDNRSVVRTVYSANMEYFKTLPVSL 1924  
Qy 1695 -----VW-----VKTKD----- 1702  
Db 1925 TALDNRAGLSPATWNFSTYELLDYLDQNRYEYIWCGLYPTTYSVNTLKNLDLGYQAK 1984  
Qy 1703 -----GSKALLATYNAAGOTNYLTNNPAEALDRI 1732  
Db 1985 PAPTAPMPKAPELDLRGHTLESAGRKIFGEVKKLOGEYKAKMAVQAVEAYGEATRRV 2044  
Qy 1733 NEQ-GIRFFHVNDGNQPPVQGRNGRIDSSASGKHSVAI--GFOAKADGEA----- 1779  
Db 2045 HDLQGRYGRKALGMDAETKEVDGIIQEFADRLRTVYAKQADQATIDAEITDKVAQRVKSQ 2104  
Qy 1780 -----AVAIQROTQAGNSIAIGDNAQATGDQSIAGTGNVNVAGKHSIGAIDPSTVKA 1832  
Db 2105 IDAVRLQAIQGRVTLAKALSALGADWRALGHSQLMQRMKDEKAGKRGAEIA----- 2157  
Qy 1833 DNSYSVGNNOFTQTDQDFGVGNNTITVESNSVALGSNSAISAGTHAGTO--AKKSDG 1890  
Db 2158 -----FYPREQT-VLAAGAGLTLS-NGAIHNGENAAQNRGPEGLKIGHATS 2204  
Qy 1891 TAGTTTATAGTGVKGFAGQTAAGVAVSG-----ASCAERRIONVAAGEV-----SAT 1938  
Db 2205 VSGSDALRDVLEKRLDIDDAALVLPNPHITFRIGAAQTSIADGAAAGPALARQARQAP 2264  
Qy 1939 STDV-----NGSOLYKATQSTANTNDELHRIQENENKANAGIS--SAMAMASMPQAYI 1991  
Db 2265 ETDMGVMDARGLSGSADALASLASDAAGLEVSRRNAQVADAGIAGPSAVAAPVAGADV 2324  
Qy 1992 PGRSMVTGGIATHNGGAVAVGLSK 2016  
Db 2325 -GVPEVTG---DOVDQPPVAVGLEQ 2345

RESULT 9  
P2P\_LACPA STANDARD; PRT; 1902 AA.  
ID Q02470;  
AC 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-  
DE ASSOCIATED SERINE PROTEINASE) (LP151).  
GN PRTP.  
OS Lactobacillus paracasei.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=1597;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCDO 151.  
RX MEDLINE=92381481; PubMed=1512565;  
RA Holck A., Naes H.;  
RT "Cloning, sequencing and expression of the gene encoding the cell-  
RT envelope-associated proteinase from Lactobacillus paracasei subsp.  
RT paracasei NCDO 151.";  
RL J. Gen. Microbiol. 138:1353-1364 (1992).  
RN [2]  
RP SEQUENCE OF 189-196.  
RX MEDLINE=92226694; PubMed=1564442;  
RA Naes H., Nissen-Meyer J.;  
RT "Purification and N-terminal amino acid sequence determination of the  
RT cell-wall-bound proteinase from Lactobacillus paracasei subsp.  
RT paracasei.";



QY 1176 FAGENG-----TTKVNGV-----VRVIGDQTKGLT-----TPKLTVCNNNGK 1214  
 Db 1196 DKAHTQALLKHFDTKVRLTAPTFTDLKFNNGSDQTSFATIKVTGTVSSDQTKVNVGDTVA- 1254  
 QY 1215 GIVIDSON-----GQNTITGLSNTLNVNTDNGKSVRTEQNGIIKDEKTRAASI 1264  
 Db 1255 --ALDAQHFSVDVPVNYGDNITK-----VTADEDEGNT-TTEQKTISSYDP----- 1299  
 QY 1265 VDLSAGFNLOQNGEAVDFVSYDT-VNFA--DGNATTAKVYDDTSKTSKVYDVND 1321  
 Db 1300 -DVLK-----NAV-----TFDOGVKFCANFNATSAKF-YDPKGTATITGKVHPT 1344  
 QY 1322 TTEVDRKDKLVK-----TFTLSTGTGANKFAL-----SNQATGDLVAKASDVAHLN 1370  
 Db 1345 TTLQVQKQKISIKNDLTFSTLQGLTGKQPGVVVGGDTTQNKTFEALFILDVA--P 1402  
 QY 1371 TSLSDIOTAKGASQANNSAGYVDADGNKVIYDSTDKYQAKNDGVPDKTEKVAOKLVA 1430  
 Db 1403 TSLSDST-----DAPVYNDPNFQITGTATDNAQVLS-----LAI 1438  
 QY 1431 QAOPPDGTLAOMNVKSVINKEQVNDANK--QGINEDNAFVKLEKASDNKTKNAAVTV 1488  
 Db 1439 NGSVASQYADININSGPKGHMAIDQPVKLEGN-----VLTVAVDSENNTTKKIIV 1493  
 QY 1489 GDLNVAQTLPTFAAGDTGTAKKLGELTLTIKGGTQTNKLTNNIGVAGTGDGFTVKLAK 1548  
 Db 1494 -----YEPKKTLLAAPTVP-----TTEPAKTVI--LTAN-----AAATGETVOYSA 1534  
 QY 1549 DLTNLNSVNAAGTKIDD--KGVSEVDSSGQAKANTPVLNSANGLDGGKVISVNGKTKD 1605  
 Db 1535 D-----GGKTYQDVPAAQVTV--VTANGTFKFSKFDLYGNESPAVDYVVTNI-----KA 1580  
 QY 1606 TDAANVOQLNEVRNLGLIGNAGNDNADGNVNIADIKKDPNSGSSSNRTVIKAGTVLGGK 1665  
 Db 1581 DDPACQQTAKAALNLT-----IASAKTSLASGKYDDATT-----TALAAA 1620  
 QY 1666 GNNDEKLTAGGIGQ--VGVDKGNANGDLSNWKTKQDSKALLATYNAAGTNYLTN 1723  
 Db 1621 ----TQKAQTAQDQDASVDSLTGANRLQ-----AINQLAAKLPAKDKTSL-- 1665  
 QY 1724 NPAAEIDRINEOGIRFPHVNDGNQPPVQGRNGIDSSASGRKHSVAIGFQAKADGEAAVAI 1783  
 Db 1666 -----NQLOSVKAALGTDLGNQDTPSTGKTFFTAALDLVA--- 1700  
 QY 1784 GRQTAQNGQSTAGNQAQATQDQSTAGTGNVNAKKGSGATGDPSTKADNSYSGVGNQ 1843  
 Db 1701 --QAQATQGT--ADQLQA-----SLAKVLDVA/LAKLABG-----IKAAATPAEYGNAK- 1743  
 QY 1844 FTDAQTQDVFYGVNNITVTESNSVALGSN-----SAISAGTHAGTQAKKSDGT 1891  
 Db 1744 --DAATGKTWYADIADTLTSGQASADADSKLAHLQAQLQSLTKVAVAEAAKTAGKDDT 1801  
 QY 1892 AGTITTAGATGTVKGFAGQATVAGVSGASGAEKRIQNVAAQEV-----SATSTD 1941  
 Db 1802 TGTSDKGGGGTTPAPAGDT-----CKDKGDEGQSPSSGGINPTKPTATTSTSD 1851  
 RESULT 10  
 OMPB\_RICJA STANDARD; PRT: 1656 AA.  
 ID OMPB\_RICJA  
 AC O06653  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
 DE [SURFACE PROTEIN ANTIGEN] (CELL SURFACE ANTIGEN) (SCAS) (ROMP B)  
 DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
 GN OMPB  
 OS Rickettsia japonica.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsia.  
 OX NCBI\_TaxID=35790;

RN SEQUENCE FROM N.A.  
 RP STRAIN=VH;  
 RC Uchiyama T.;  
 RA "Sequencing of the gene encoding the protein rOmp B of Rickettsia  
 RT japonica";  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
 CC SIMILARITY).  
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL: AB003681; BAA20138.1;  
 DR Antigen: S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.  
 KW CHAIN 1 1338 32 KDA BETA PEPTIDE.  
 FT CHAIN 1339 1656 POLY-GLY.  
 FT DOMAIN 528 533  
 FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9D5999F CRC64;  
 SQ  
 Query Match 4.3%; Score 443.5; DB 1; Length 1656;  
 Best local similarity 21.4%; Pred. No. 4.1e-09;  
 Matches 439; Conservative 237; Mismatches 750; Indels 623; Gaps 106;  
 QY 191 SSKDNVKKYRRRTTASGHASTAVGAMSYAQHIESNAGFTRATKASVLAAGLAETAEQS 250  
 Db 2 AQPFLKLLISAGLVTAATSTAVSPAGSAGAAIQONRTN-----GVATTVDG-- 52  
 QY 251 TTAIGSDATSSSLGALACAGTCACTRAQLOGSIALGCVSVTOSDNNRKPATPTNOALDPKF 310  
 Db 53 ---VGFDQT-----VAL-----ANVAVAPNAVITANANG----- 79  
 QY 311 QATNNTKAGPLSGISIKRKIIINVAGYKNTDVAVVAOLEAVKWKERRITF--QDSD 368  
 Db 80 -INLNTPAGEFN-----GLFLSNANLA-----VTVSEDTTLGFINNAAN 118  
 QY 369 NSTDVKIGLD--NTLTIKGAETN---ALTDNNIGVYKADNSGLKVLAKTLNLTENV 423  
 Db 119 NANRPNLTLDAGKTLTTTGCGITNVOSAAHNAQNLIVAKF--NGGAAI---ANNDLSGLG 173  
 QY 424 TTTLNATTTTVKVGSSSSSTTAE---LLSDSLTPTQPTNG-----SOSTSKTVYG 468  
 Db 174 TIDFGAASTLVFDLANPTTQKAPLILADNALVINGANGTLNVNTNGFIQVSKSEATVKA 233  
 QY 469 VN-----GVKFTNNAETTAAGITRITRDKIGFARGDGVDEKQAPYLDKKQLKVGSAIT 523  
 Db 234 INIGDGGGFMFTNATNANAL-----NQAGGTTIN 264  
 QY 524 IDNGIDAGNKKISNLAKGSSANDAVTIEQLKAAKPTLNAGAGISVTPTEISYDAKSGNVT 583  
 Db 265 F-NGTD--GTGRVLVLLSKNGAATD-----FNVTSGLSGNLKGI--IELNTVAINGQLI 312  
 QY 584 APTYNIGVKYTELNSDGTSDKFSVKGSGTNNLSLVTAHLASYLNEVNRNADSLAQSTFKV 643  
 Db 313 A---NAGPANAVI-----GTNNG---AGRAAGFVSV----- 338  
 QY 644 EEDDDANAIVT-----AKDTTKNAGAVSILKLGKNGLT-----VATKKDGTVTTF 689  
 Db 339 ----DNKGAATIDGOVYAKD-----MWIQSANANGVNFHRLVDVGIDGTGTAFTAA 386  
 QY 690 ---GLSQDSGLTIGKSTLNN---DGLTVKDTNEQIQVANGIKFTNVNGSNPGTGIANTAR 744

Db 397 SIVALTQNS--NFGYTDGPNLAAQVTVDPD---MTLTCN---FTG-DANNPG-----NTAG 433  
Qy 745 ITRDKIGFAGSDGAVDTNKPQLDQKLVGNVKTIN--TGINAGGKAITGLSPILPSIAD 802  
Db 434 V-----ITFAANGTLASASAD-----ANVAYTNNTAIEASGVGVQLSGT-----474  
Qy 803 QSSRNIELGNT-----TODKDSNAASINDILNTGPNLKNNNPIDFVSTYDVIDVFANGN 857  
Db 475 -HTAELRLGNAGSVFKLADGVFIN-GKVNTQTVLVGGVLAAGAITLDGSAIT-TGDIINGG 531  
Qy 858 ATTATVTHDTANKTSKVYVDVNDVDTIHLTGCTDDNKKLVKTKLANKTSANGNTATNFN 917  
Db 532 GGAALQSLTLANDATK-----TLTLGGA-----NIIISANGGT-INFQ 567  
Qy 918 VNSSDEDAVNKADIAENLTLAKEIHTTKGTADTALQTFYVKVDENNNADDAITAVG 977  
Db 568 ANG-----CTIKILTSTQNNIVVD-----CDLAIAI-----DQTVGVVDASSLTNA 606  
Qy 978 OKNANNQVNTLTLKGENGLNKTDKNGTVTFGI-----NTTSG-----LKAGKSTLNDGGLSI 1030  
Db 607 Q-----TLTISG-----TIGLIICANNITLQGFNIGSKTTLNGGNVAI 644  
Qy 1031 KNPTGSEIOVGADS-VKFAK-----VNNNGVVCAGID-GTTR 1066  
Db 645 -----NELVIGNNGSVQFAHTYLTITRTNAAGOKLIIFPNVNNNTTLAGATNLGSA 698  
Qy 1067 ITRDEIGFTGNG-----SLDKSPHLSKDGINAGKKTITNIQSGEI 1108  
Db 699 NPLAEINF-GSKGARADTVLVGEGVNLATNIITDANVGSVFVFNAGK---NIVSGTV 754  
Qy 1109 AQNSHDVATGKGIYDLKTELENKISSAKTAQNSLHEFSVADEOGNPTVSNPSYDTS 1168  
Db 755 -----GG-----QQGNKENTVALDNGTTVKFLGNATFNNGNTTIAANSTLQISGN 798  
Qy 1169 KTSVITPAGENGITTKVNGKVVVRGID-QPKGLTTPKLTGVNNGKGIIVDSONGQTI 1227  
Db 799 YTADEFIASADGTGIVEFVNTGPIVNTLKNQAPVNALQIITVSGFNVVNGEIGNAGNYH 858  
Qy 1228 TGLSNTLANVTDNKSVRTTEBQNIIDKEDTRAASIVDVLISAGFNLGNGEAVDF-----1283  
Db 859 GAMTDTIAFENSSILGAVLFLPSPFNENDAGNTIPLTKSTV-----GNETAECFSVPS 911  
Qy 1284 -VSTYDVTNFADG-----NATTAKVYDDTSKTSKVYVDVNV 1319  
Db 912 VIVSGVDSV-IADGQVIGDQNNIVGLGLSDNGIIVNATTLIYAG-----IGTINN 960  
Qy 1320 DDTTIEVKDKKLVKVTTLT-----STGFGANKFALSNOATGDALVKASDIAVHLNLTSGD 1375  
Db 961 NQGVITLISG---GVNPTGTVYGLTGIGASKFKQVTFETD--YNNLGNIIATNTTINDG 1015  
Qy 1376 IQTAGASQANNSAGYVDADG-----NKVIYDSTDNKYIOAKNDGTVDK 1419  
Db 1016 VTVITGGIAGGIAG-TDFDCKITLGSVNGNANVRFADGIFSNSTSMIVTKKANNGTVTY 1074  
Qy 1420 TKEVAKDKLVAQAOTP-----DGTLAQMNVKSVINKBQVNDANKQGINEDNAFVKGLE 1473  
Db 1075 LGNAFVGN-IGDSOTPVASVRFSTGNSNAGLKNYISQVIDFGTYNLGVNSNVLGGST 1133  
Qy 1474 KAASDNKTNAATVVDLNAVAQTPPLIFAGDTGTAK--KLGETILTIGKQTDINKLTDN 1531  
Db 1134 TAIN-----GKIDLLTNT-LTFAGGTSTWGNNTSTETTLTLANG-----1171  
Qy 1532 NIGVAGTGDGFTVKLAKDLTLNLSNVNAGGTGI---DKGVSVFVDS-----SGOAKANTP 1582  
Db 1172 NIGHIVIAEG-----AQVNATTGTTTINVDNANANFSGTOTYTLLOGGARFNGT 1222  
Qy 1583 VLSANGDLGKGVTSNKGKTKDTDAAN-----VOOLNEVRNLL--GLGNAGNDAGNQV 1636  
Db 1223 LGGNFNTVTSNRNFVNYGL-----IRAAQNDYVITRTNAENAVITNDITNSFPFGAPGVGQ 1278  
Qy 1637 NIADIKKDPNSGSSNRNVIK-----AGTVLGGKGNNDTEKLTATGCIQGVGDKD---1685

Db 1279 NVTFVNTATNTAAVNNLLAKNSADSANFVGTV-----TDTSAAITNA-QLDVAKDIOA 1332  
Qy 1686 --GNANGDLSNVVWTKQKDGSKKALLATYNAAQGTNYLTNNPAEAIDRINEQGI--REFH 1741  
Db 1333 QLGNRGLGALRYLGT-PEWVGSEAGAI PAAVAAGD-----EAVDNV-AYGIWAKPFY 1381  
Qy 1742 VNDNQEPVVOGRNGIDSSAGSKHSVAIFGFOAKADGEAAVAIGROTQAGNQ---SIAIGD 1798  
Db 1382 T-----DAHQSCKGGLA-GYRAKTTG---IVIGLDTLANNLMIGAIGI 1422  
Qy 1799 NAQTGQDSQIAIGTGNVVAGKHSAGIDPSTVKADNSYSVGN---NOFTDATOTDVF 1854  
Db 1423 TKTDIKHODYKKDKTDVNGFSFSLYGAQFV--ENFFAQGSAIFSLNQVKNKSRYFFD 1480  
Qy 1855 VGNNTVTSNSVALGNSAISAGTH-----AG-TQAKKSDCTAGTT 1895  
Db 1481 ANGNM---SKQIAAGNYDNMTFGNLTGVGDYDYNAMQGVLTPTMAGLSYKSSDENYKRT 1536  
Qy 1896 TTAGATGTVKG-FAGOTAVGAVSGVSGAERRIQNVAAGEVSATS---TDAVNGSQLYKA 1951  
Db 1537 GTTVANKQVNSKFSVRT---DLIVGA-----KVAGGTMTNITDFAVYPEVHAFVHKV 1585  
Qy 1952 TOSTANATNELDHR---IHONENKA---NAGISSAMAMASMPQAYIPGRSMVTGGIAT 2003  
Db 1586 TGRUSKTQSVLDGQVTPCISQPDRTAKTSYNLGLSASIRSDAKMEYIGYDAQIASKYTA 1645  
Qy 2004 HNGOGAVAV 2012  
Db 1646 HOGTLKVRV 1654

RESULT 11  
PIP\_LACLC  
ID PIP\_LACLC STANDARD; PRT; 1902 AA.  
AC P16271;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE PI-TYPE PROTEINASE PRECURSOR (EC 3.4.21.-) (WALL-ASSOCIATED SERINE  
DE DE PROTEINASE).  
GN PRTP.  
OG Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
OC Plasmid pMW05.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Lactococcus.  
OX NCBI\_TaxID=1359;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=WG2;  
RX MEDLINE=88149035; PubMed=3278687;  
RA Kok J., Leenhouts K.J., Haandrikman A.J., Ledeboer A.M., Venema G.;  
RT "Nucleotide sequence of the cell wall proteinase gene of  
RT Streptococcus cremoris Wg2.";  
RL Appl. Environ. Microbiol. 54:231-238(1988).  
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
CC -!- GROWTH OF THE BACTERIA ON MILK.  
CC -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD  
CC SPECIFICITY, ALTHOUGH SOME SUBSTRATE PREFERENCE HAVE BEEN NOTED.  
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND  
CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,  
CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED  
CC INSULIN B-CHAIN.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
CC SUBTILASE FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL: M24767; AAAL7677.1; -.
DR HSP: O99405; 1MPT.
DR MEROPS: S08_019; -.
DR InterPro: IPR000209; -.
DR pfam: PF00746; Gram_pos_anchor; 1.
DR pfam: PF00082; Peptidase_S8; 3.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
DR Hydrolase: Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1875
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
FT SEQUENCE 1902 AA; 199910 MW; 2901C7F19B2E5D0B CRC64;

Query Match 4.3%; Score 441.5; DB 1; Length 1902;
Best Local Similarity 20.7%; Pred. No. 5.7e-09;
Matches 459; Conservative 238; Mismatches 829; Indels 693; Gaps 97;

QY 40 SVCTLSFARIAALAVLIGATLCSAYAKKDKTHIAIGEQNPRRSGTAKADGDAIAI 99
DB 9 SILLAGTVAGALAVLPVGEIOAKAAISQTKGSSLA-----NTVPAATAKQ-----AA 57
QY 100 GENAAGGQAATAGSSNKTNGSSLDKI-----GTDATGQES- 137
DB 58 TDTTAATTQALATQALAKGIDYNKLNKVKQQDIYVDIVQMSAAPASENGTLRTDYSST 117
QY 138 -----TAIGDVKVA-----SGDA-----SIAIGSDDLHLHDGHPNK- 169
DB 118 AEIQOETNKVIAAQAASVAAVEQVTOOTAGESYGVVNGFSTKRVVDIPKLAQIAGVKT 177
QY 170 -----HPKGLINDLINGHAVL-----KEIRSSKDNVDKY 199
DB 178 VTLAKVPTPTDAKASMANVQAVMSNYKKGEGTVSVSDIGIDPTHKDMRLSDDKDKVL 237
QY 200 RRTTASGHASTAVG-----AMSYAQG-----HFSNAFGTRATAKS 234
DB 238 TKSDVEKFTDTAKHGRYFNKSPYGFENYADNNDDTTTDDTDEQHGHWAGIIGANGTGDD 297
QY 235 AVSLAVGLAATAEGOSTIAIGSDATSSSIALGALGAGTQAOLGSGSVVYQSDNN 294
DB 298 PAKSVGVGAPEAQLLAKMVFYNSDTSATTSGLVSA-----IEDSAKIGADVL-----NM 348
QY 295 SRPAPTPTQALDPKFPATNNTK-----AGPLSIGNSIKRKRIINVGAGVKNKTDAVNAQL 350
DB 349 SLGSDSGNQTLEDPELAAYQONANESGTAAVISAGNSGTSG-----SATEGVNK--DIYGLQDN 404
QY 351 EAVVKWAKERRITFOGDDNSTDV-----KIGLDNLTIAIKGAETNALTDNNI 397
DB 405 EMVGTGCTRGATTVASAENTDVIQAVTITDGTGLGLG-PGTIQLSSNDFGSPDQKFF 463
QY 398 GVVKE-----ADNSGLKVLKAK-----TLNLT 420
DB 464 YVKDASGNLSKALADYTADAKG-KIAIVKRGELSFDDKQYAAAGAAAGLIIVNNDDGT 522
QY 421 EVNTTILNATTIVKVGSSSSTTAELSD-----SLTFTQPTNGSQSTSK 464
DB 523 ATPVTSMALTTFTTFCGLSVTCQKLVDMVTAHPDDSLGLVKIALTLVPNQKYTEDKMSDF 582

```

```

QY 465 TVYG-VNGVKFTNNAETTAAGTTRITRDKIGFARDGDVDEKQAPYLDKKQLKVGSVAIT 523
DB 583 TSYGVPVNSLSE--KPDITAPGGNIWSTONNNGTNSGT-SMASPFIAQSQAALLQALNN 639
QY 524 IDNGIDAGNKKISNLAKGSSANDAVTIEQLKAAKP-----TLNAGAGISVTP 571
DB 640 KNPFPYAYKQL-----KGTALTDFLKTVMEMTAQPIINDINNVIVISPRQAGL--VDV 693
QY 572 EISYDAKSGNVAPTNYNIGVKTTELNSDGTSDKFSVKGSGTNNNSLVTAELHASYLNEVNR 631
DB 694 KAAIDALEKNPSTVVAENGYPVAVELKDFSTDK-TFKLTFTN-----NGLTVATKKDGT 734
QY 632 TADSALOSFTVKEEDDDANAITVAKOTTKNAGAVSILKLGK-----NGLTVATKKDGT 666
DB 735 ---STHETLYQMSDNTDITNAVITSA-TDPNSGVLYDKKIDGAIAKAGSNITVPAGTAQ 790
QY 687 VTEGLSQDSGLTIGKSTLNDGLVAKDTNEQIQVANGIKFTNVNGSNPGTGITANTARIT 746
DB 791 IEFITLS-----LPKSPDQOFVGS-----PLNFKGS-----DGSRLN 822
QY 747 RDKIGFAG--SDGAVDTNKPVDQDKLOVGNVKITNTGINAGKAITGLSPTLPSIADQS 804
DB 823 LPYMGFFGDWNGKI-----VD---SLNGITYSPAGNFG-----TVPLLTKN 863
QY 805 SRNIELGNTIQDKDKSNA-----ASINDILNTGFNLKNNNNNPIDFVSTYDIV 851
DB 864 TGTQYGGWTDADGNQTVDOQATAFSSDKNALYNDISKYLLRNISN-----VQVDIL 918
QY 852 DPANGN-----ATTATVTHDTANKTSKVYDVNVDDTTIHLTGTDNKKLGKVTTKLN 904
DB 919 D-GQGNKVTTLSSSTNLTKTYNAHQOQYIVYNAPAWDGTIV-----DQDGNK- 967
QY 905 KTSANGNTATNPNVNSSDEDALVNAKDIAENLTLAKEIHTTKGTADTALOTFTVKKYDE 964
DB 968 -----TADGSGTYRISGYPE 983
QY 965 NNNAD-----DANAITVGQKNANNQVNTLTLLKGNG-----LNITDKNG----- 1004
DB 984 GGDKRQVDVFPKLDKSKAPTVRHV-----ALSAKTENGKTQYLYLTAEAKDDLGLD 1036
QY 1005 -TVTFGINTSGLKA---CKSTLNDGGLSIKNPTGSEQIQVGCADGVKFAKYNNGVWVG 1060
DB 1037 KSVKTAINEVTNLDAFTDAGTADGTYKIEPLSDEQAQALGNDGNSAEL----- 1087
QY 1061 IDGTTTRITDEIGFTGNSLSDKSPHLSKDGINAGG--KKTINOSGEIAQNSHDVIT- 1117
DB 1088 -----YLTDNASNATDODASVQKPGSTFELIVNGGGIPDKISSTTG-----YEANTQ 1136
QY 1118 GGIYDLKTELENKISSTAKTAQNSLHEFSVA-DEOGNNTVSNPNYSSYDTSKTSDVITF 1176
DB 1137 GGGTYTFSCTYPAADVGTYTNAQGGKHIDLTNYDAATNSFTASMPVYNADYAAQVDLYAD 1196
QY 1177 AGENGI-----TTKVNKGV-----VRVGIDQTKGLT-----TPKLTGVNNGNGK 1215
DB 1197 KAHTQLLKHPDTKVRLTAPTFTDLKFNNGSDQTSSEATIKVTGTVSADTKTVNVGDVFA-- 1254
QY 1216 IVIDSQN-----GQNTITGLSNTLANVTNNDKSGVYRTTEOGNIKIDEDKTRAASIV 1265
DB 1255 -ALDAQHFSVDVPNVYGDNTIK-----VTATDEGNT-TTEQKITSYD- 1299
QY 1266 DVLSAGFNLOGNEAVDFVSTYDTVNFADGNATTAKVYDDTTSKTSKVYDVNVDDTTIE 1325
DB 1300 DMLK-----NSVTFDQGVTFGANEF---NATSAKF-YDPKGTIATITGKVKHPTTLQ 1348
QY 1326 VKDKKLGVK-----TTTLTSTGTGANKFAL-----SNQATGALYKASDIVAHLTSLG 1374
DB 1349 VDGQIPIKIDDLTFSFTLDLGLTGOKPFGVVVGDTTQNTKTFQEBALTFILDAVA--PTLSL 1406
QY 1375 DIQTAKGASQANNAGYVDADGNKVIYDSDTN-KYIOAKNDGT-----VDKTRKVAOK 1427
DB 1407 ESST-----DAPVTNDPNFOITGTATDNQYLSLSINGSSVASQYVDININSKPG 1458
QY 1428 LVAQAQTPDGTLAQMNKSVINKPEQVNDANKKOGINEDNAFVKGLEKAASDNKTKNAVT 1487

```

Db 1459 HMAIDQ-PVKLEKGNLVV-----AVTD-----SEDTTKNITVYEPKKT-L-APT 1505  
Qy 1488 VGDNLNAVQPTPLTADGTGTTAKKLGELT-LT-IGGQTDNTKLTNDNIGVAGTGTGFTVCLA 1547  
Db 1506 V-----TP-----SITEPAKTVTLTANSAT-----GETVOYS 1533  
Qy 1548 KDLNLNSVNAAGTKIDD---KGVSVFVSSGOAKANTPVLSANGLDLGGKVISVWVGKTK 1604  
Db 1534 AD-----GKTYQDVPAAGVT-VTANGTFKFKSTDLYGNSPVDVYVVTNI-----K 1579  
Qy 1605 DTDAAVQQL-NEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSSNRTVIKAGTVLG 1663  
Db 1580 ADPAQQAQKQELTNL-----IASAKTLSAGKVDAT-----TALA 1618  
Qy 1664 GKGNDTEKLATGIGQ--VGVKDKNANGDLSNVWVTKQDKGKALLATVYNAAGQTNYL 1721  
Db 1619 AA-----TQKAQTALDQTNASVDSLGTANRDLQT-----AINQLAAKLPAKDKTSL 1665  
Qy 1722 TNNPAEALDRINEGIRFFHVNDGNQEPVVOGRNGIDSSAGKHSVAIGFOAKADGAAV 1781  
Db 1666 -----NOLQSVKAALGTDLGNQTDPSGTGKTFPTAALDDLVA- 1700  
Qy 1782 AIGRQTAGNOSIAIGNAATGDSIAIGTGNVVGKHSAGIADPSTVKADNSYSVGN 1841  
Db 1701 ---QAQGTQT---DQHQAT---LAKVLDVLAKEG-----IKAATPAEVGNA 1742  
Qy 1842 NQFTDATQTDVFGVGNITVTESNVALGSN-----SAISAGTHAGTQAKKSD 1889  
Db 1743 K---DAATGKTYADIADTLTSGQASADSKLAHLQALQSLKTKVAAYEAATVKG 1799  
Qy 1890 GTAGTTTAGATVKGAGTAVGAVSVGASGAERRIQNVAGEV-----SATSTD 1941  
Db 1800 GTTCTSDKGGQGPAPAPG-----DICKDKGDEGSPSSGGNIPTNPATTTSTSD 1851

RESULT 12

P3P\_LACLC STANDARD; PRT; 1902 AA.  
ID P3P\_LACLC  
AC P15292;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE PIII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-  
DE ASSOCIATED SERINE PROTEINASE).  
GN PRTP.  
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
OG Plasmid.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Lactococcus.  
OX NCBI\_TaxID=1359;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.  
RC STRAIN-SK11.  
RX MEDLINE-89340435; PubMed-2760036;  
RA Vos P., Simons G., Siezen R.J., de Vos W.M.;  
RT "Primary structure and organization of the gene for a procaryotic,  
cell envelope-located serine proteinase."  
RL J. Biol. Chem. 264:13579-13585(1989).  
CC -I- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
CC GROWTH OF THE BACTERIA ON MILK.  
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.  
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
CC SUBTILASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

DR EMBL; J04962; AAA03533.1; ALT\_SEQ.  
DR PIR; A32634; A32634.  
DR HSP; P00782; 28BT.  
DR MEROPS; S08.019; .  
DR InterPro; IPR000209; .  
DR InterPro; IPR001899; .  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF00082; Peptidase\_S8; 3.  
DR PRINTS; PR00723; SUBTILISIN.  
DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; 1.  
KW Hydrolase; Serine protease; Cell wall; zymogen; signal; Plasmid;  
KW Transmembrane.  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT DOMAIN 1867 1872  
SQ SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;

Query Match 4.2%; Score 430.5; DB 1; Length 1902;  
Best Local Similarity 20.5%; Pred. No. 1.4e-08;  
Matches 452; Conservative 241; Mismatches 838; Indels 679; Gaps 95;

Qy 40 SVCTLSFARIAALAVLVIGATLSGSAYAAKKDKTHIAIGEQNPPRRSGTAKADGDRATAI 99  
Db 9 SILLAGTVALGALAVLPVGEIQAKAAISQTKGSLA-----NVTAAFAKQ-----AA 57  
Qy 100 GENAAGGQQAIAIGSSNKTNGVSSLDKI--GT 130  
Db 58 TDTTAATTNQAIALQAAGKIDYNKLNKVKVQQDIDVDVIVQMSAAPASENGILRTDYSST 117  
Qy 131 DATQES---IAIGDVKA-----SGDA-----SIAIGSDDLHLDDQHNPK- 169  
Db 118 AEIOQETNKVIAQAQAKAAVEQVTTQTAGESYGVYVNGFSTKVRVVDIPKLQIAGVKT 177  
Qy 170 -----HPKCTLINDLINGHAVL-----KEIRSSKNDVKY 199  
Db 178 VTLAKVYPTDAKANSMANVQAVNSYKKGEGIVVSVDSGIDPTHKDMRLSDDKDKVL 237  
Qy 200 RRTTASGHASTAVG-----AMSYAQG-----HFSNAGFGRATAKS 234  
Db 238 TKSDVEKFTDVTVKHGRYFNSKVPYGFVADNNDITDDKVDQEHGMHVAGIANGTGDD 297  
Qy 235 AYSLAVGLAATAEQGSTIAIGSDATSSLSGAIAGAGTGAQLOGSIALGQGSVVTQSDNN 294  
Db 298 PAKSVVGVAPQAQLLAKMVKFNSDTSKATGS-----ATVVSATIEDSAKIGADV 348  
Qy 295 SRPVTPTNQALDPKFOATNTK-----AGPLSIGNSIKRKLIINVGAVNKTDVAVNAQL 350  
Db 349 SLGNSNGQTLDEPELAQVQANESGTAAVISAGNSGTSG---SATEGVNK-DYIGLQDN 404  
Qy 351 EAVYKWKAKERRITFGDDNSTDVKGIDNLTITL----- 383  
Db 405 EMVSGPGTSRCATTVASAENTDV---ITQAVTITDGTGLQLGPETIQLSHDFTGSFQK 461  
Qy 384 -----KGAETNALT-----NNIGVYKKEADNS-GLKVKLAKTLN--NLFEVNT-- 424  
Db 462 KFYIVKDAQSNLSKALADYTADAKGTAIVKRGESFDDKQKYAQAAGAGLIIVTDG 521  
Qy 425 -----TTLNATTTVKVGSSSTTALLSDSLTF-----TOPNTGSSQSTS 463  
Db 522 TATPWTISALITTTPTFTGLSSVVGKLVWVTAHPDDSLGSKITLMLPNOKYTEDKMSD 581

QY 464 KTVYG-VNGVKFTNAETAAITGTRITRDKIGFARGDGVDEKQAPYLDKKQKLVGSVAI 522  
 Db 582 FTSYGPVSNLSE--KPDITAPCGNIWSTQNNNGYTNMSGT-SMASPFITAGSOALLKQALN 638  
 QY 523 TIDNGIDAGNKKISNLAKSSANDAVTTEOLKAAP-----TLNAGAGISVTP 570  
 Db 639 KNNPPYAYKQL-----KGTALTDFELKTVMETAQPIINDINYNVIVSPRGGAGL--VD 692  
 QY 571 TEISVDAKSONVTAPTNYTNGVKTTELSNDSCTSKFVSXGSGTNNLSLVTAELHASYLNEVN 630  
 Db 693 VKAAIDALEKNPSTVVAENGYPAVELKDFSTDK-----TEKLTFTN 734  
 QY 631 RTADSLAQSEETKEEDDDANAITVAKDTTKNAGAVSILKLGK-----NGLTVATKDKG 685  
 Db 735 RTT-----HELTYQMSNTDTNAVYTSA--TDPNSGVLYDKKIDGAAKAGSNITVPAGTA 789  
 QY 686 TVTFGLSQDSGLTIGKSTLNNDLGTVKDTNEQIQVGANGIKFTNYNGSNPGTGIANATARI 745  
 Db 790 QIETLS-----LPKSFQOQFVEG-----FLNFKGS-----DGSRL 821  
 QY 746 TRDKIGFAG--SPGAVDTNPKYLDQDKLQGVNVTNTGINAGGKAITCLSLPLPSIADQ 803  
 Db 822 NLPYMGFGWINDGKI-----VD-----SLNGITYSPAGNFG-----TVPLLNK 862  
 QY 804 SSRNIEGNTIQDKKNSA-----ASINDILMTGFNLKNNNNPIDFVSTYDI 850  
 Db 863 NTGTQYGGWTDADGKNTVDDQAIASFSSDKNLYNDISMKYLLRNISN-----VOVDI 917  
 QY 851 VDFANGNATTAT-----VHTDANKTSKVYDVNVDVDTIHLTGDDNKKLGKVTIKLN 904  
 Db 918 LDGQGNKVTLLSSSTNRKTKYNAHSQYIYYPAPAWDFTY-----DORDGNIK 967  
 QY 905 KTSANGTATNFVNDSDEALVNAKDIAENLTLAKEIHTTKGTADTALOTFTVKKVDE 964  
 Db 968 -----TADGGSYTYRISGVE 983  
 QY 965 NNNAD-----DANAITYGQKANNQVNTLTKGNG-----LNKTKNG-----1004  
 Db 984 GGDKQRFVDFPKLDSKAPTVRHV-----ALSAPKENGKTQYILTAEAKDDLGLDAT 1036  
 QY 1005 -TVTFGINNTSGLKA--GKSTLNDGGLSTKNPTGSEQIOVGADGVKFAKVNNGVVGAG 1060  
 Db 1037 KSVKTEINEVNTLDAFTFDAGTTADGYTKIETPLSDQALGNGDSAEI-----1087  
 QY 1061 IDGTRITRDEIGFTGTNGSDKSPHLSDGKINAGG--KAITNIOSEIAQNSHDAVT--1117  
 Db 1088 -----YLTDNASNATDQASVQKPGSTFOLIVNGGGTPDKISSTTGT-----YEANTQ 1136  
 QY 1118 GKKIYDLKTELENKISSHTAKTQNSLHEFSVA--DEOGNNTVSNPYSSYDTSKTSDVITE 1176  
 Db 1137 GGGTYTFSGTYPAADVGYITDQAGKKHDLNTTYDAATNSFTASMPVTNADYAAQVDLYAD 1196  
 QY 1177 ACENG1-----TPKVNKGV-----VRVGIDQTKGLT-----TPKLTGVNNNGK 1215  
 Db 1197 KAHTQLLKHFDKVRMLAPFTFDLKNFNGSDQTSKATKVTGVSADTKTVNNGHTVA--1254  
 QY 1216 IVIDSONQNTITGLS-----NTLANVTNDKGSVRTTEQCNIIKDEKTPRAASIVDLSAGF 1272  
 Db 1255 -ALDAQHFSVDVPVNYGDNITKVTATDKDGNNTTTEQKTTTSSYDP-----DMLKKS 1306  
 QY 1273 NLOGNEADVFTSYDT-VNEADG--NATKAVTYDDTSKTSKVYDVYVNDVDTTIEYKDK 1329  
 Db 1307 -----TFDQGVKFGTNKFNATSAKF-YDPKTGTATITGKVKHPTTLTQVLDGK 1352  
 QY 1330 KLGK-----TTTTSTGTGANKFAL-----SNQATGDALVKASDIYVAHLNLTSGDIQT 1378  
 Db 1353 QIPKDDLTFSFTLDLGLQKPFVGVVGDITQNTKTFEALSFIIDAVA--PTLSLDSST 1410  
 QY 1379 AKGASQANSAGYVADGKNKYVDSTDNKYXOAKNDGTVDKTKKVAQKLVAAQATPDGT 1438  
 Db 1411 -----DAPVVTNPNFQITGATDINAQYLS-----1435  
 QY 1439 LAQMNKSVINK-EQVNDANKKG-----INEDNAFVKG-----LEKAASDNKTKNAVTVGDL 1491

Db 1436 -LSINGSSVASOVEDININSKPGHMAIDQPKLVLEGNVLTVAVTDSDN---TTTKNI 1491  
 QY 1492 NAVAQPLTIFAGDTGTAKKLGETLTKGGQDTNKLTDNNIGVAGTDFTVKLAKDLT 1551  
 Db 1492 TVYEFKTLAAPTIVPS-----TTEPAQIVT--LTAN-----AAATGETVQVSAD--1535  
 QY 1552 NLNSVNAAGTKIDD---KGVSVFVDSGQAKANTPVLNSANGLDLGGKVISNVGKGTQDTA 1608  
 Db 1536 -----GGKTYQDVPAAGVT-ITANGTEKFKSTDLYGNESPAVDYVVTNI---KADDP 1583  
 QY 1609 ANVOOL-NEVRNLLGLGNAGNDNADGNQVNIADIKDPDNGSGSSNRTVIKAGTVLGGKGN 1667  
 Db 1584 AOLQAARQELTNL-----IASAKTLSASGKYDDATT---TALAAA--1620  
 QY 1668 NDTKELATGGIQ--VGVDKDGNDGSLNSVWVTKQDKGSKALLATYNAAGOTNYLTNNP 1725  
 Db 1621 --TOKAOTALDQTNASVDSLGTANRDLQ-----AINQLAAKLPAADKKTSL--1665  
 QY 1726 AEATRINEOGIRFFHYNDGNQEPVVGRRNGIDSSASGKHSVAIGFOAKADGEAAVAIGR 1785  
 Db 1666 -----NQLSVKDALGTLGNQTDPTSTGKTFTTAAALDDLVA-----1700  
 QY 1786 QTOAGNOSIAIGDNAQATGDQSIAGTGNVNVAGKHSAGIDPSTVKADNSYSVGNNNQFT 1845  
 Db 1701 QAQAGTQI---DDQLQAI---LAKILDEVLAKEG-----IKAATPAEVGNAR--1743  
 QY 1846 DATQDTDFGVGNNTVTFESNSVALGNSN-----SAISAGTHAGTQAKSKDGTAG 1893  
 Db 1744 DAATGKTWYADTADTLTSGQASADASDKLAHLQALQSLKTKVAAVAEAAKTVGKGDTTG 1803  
 QY 1894 TTTTATAGTGVKGFAGTAVGAVSVG---ASGAERRIIONVAAGEVSATST 1940  
 Db 1804 TSDKGGGOGTAPAPGDTGDKDGGQSPSSGG-----NIPKPTATTST 1848  
 RESULT 13  
 ID HLVA\_SERMA STANDARD; PRG: 1608 AA.  
 AC P15320;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-NOV-1990 (Rel. 16, Last annotation update)  
 DE HEMOLYSIN PRECURSOR.  
 GN SHLA.  
 OS Serratia marcescens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Serratia.  
 OX NCBI\_TaxID=615;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.  
 RC STRAIN=SN8;  
 RX MEDLINE=88257037; PubMed=3290200;  
 RA Poole K., Schlebel E., Braun V.;  
 RT "Molecular characterization of the hemolysin determinant of Serratia  
 marcescens";  
 RL J. Bacteriol. 170:3177-3188(1988).  
 CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD  
 CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY  
 DEFINED.  
 CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM  
 ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA  
 REQUIRES SHLB FUNCTION.  
 CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
 CC -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).





FT	CONFLICT	1317	1317	M -> S (IN REF. 3)	
SEQ	SEQUENCE	1325 AA;	136514 MW;	26A3A06FA19AD7D CRC64;	
	Query Match	3.6%;	Score 375;	DB 1;	Length 1325;
	Best Local Similarity	23.7%;	Pred. No. 9e-07;		
	Matches	346;	Conservative 175;	Mismatches 578;	Indels 360;
					Gaps 83;
QY	658	DTTKNAGAVSIILKKGKNGLVATKKGDTVTYFG--LSQDSGLTTCKS--TLNNDGLVAKVDIN 715			
DB	21	ELTRRAGKSTVNLKRSGLTT--KFSRLTGLVLLASGSAGSLVDNDQITNIDTD 77			
QY	716	--EQIQGANGIKETNYGNSPGGIANTARTIRDIGF--AGSDGAVD----- 760			
DB	78	VAYDAYLVGWTGVNLILAG---GNASLTITITSVIGANEDSECTVNLGCTWRLYDS 133			
QY	761	-TNKPDLDDKQLQGVNKTITNTGINAGG---KAITGLSPTLPISADQSRNIEL----- 810			
DB	134	GNNARPLNVGQSGTGLNKKQGHVDGGLYRLGSGTGGVTVNVEGDSVLTTFELFEGS 193			
QY	811	--GNTTODKQSN--AASINDIL--NTGNN-----LKNNNPDPFVSTYDIDVF 853			
DB	194	YGTCSLNIITDKGVVTSIVAILGYAGSNGQVVVEKGGEWLIKNDSSIEF-----QIQNQ 249			
QY	854	ANGNAT-----TATVTHDTANKTSKVYDVYNDVDTTHLTGTDGDNKKGVTTKLNKT 906			
DB	250	GTGEATIREGLVTAENTIIIGNATGICGLNVQDDSVI-----TVRRLYN 295			
QY	907	SANGNTATPNVN---SSDEDALVNKADIAENLTLAKEIH--TTTGTADTALQTFVKKV 962			
DB	296	GYFGNGTVNISNGLINNKEYSLVGQDGSBGVWVTDKGHWNEFLGTGEAPRIYIIGDAG 355			
QY	963	DENNAD-----DANATTVGKNANNOVNTLTKEGINKITDKNGTVFGINTTSLGK 1017			
DB	356	DDELNVSEKGVDSGIITAGKET-----GTGNITVK-DKNSVIT-NLCTNLGYD 403			
QY	1018	A-GRSTLNDGSLSKNPTGSEOTGVGADGV-----KFAKVNNN-----GVVGAG--- 1060			
DB	404	GHGEMNISNOGLVSVN--GGSLGYGETGVGNVSIITGMMMEVKNKNVYTTIGVAGVGNLN 461			
QY	1061	IDGTRITRDEIGTGN---GSLDKSPHLSKD--GINAG--GKKTINQSQGEIAQNS 1112			
DB	462	ISDGKGFVSQNIITFLGDKASIGTILNMDATSSFDTGVINGVNGFSGGVNVSNGATLNST 521			
QY	1113	HDVATVGGK-----IYDLKTELENKISSTAKTAQNSLHEFSVADEGNNFTVSNPSYVD 1167			
DB	522	GYGFIGGNASCKGIVNISTDSLNLKTSSTNAO--LLQGVGLGTGELNITTTGGIVKARD 579			
QY	1168	SKTSVDITFACENGITTKVNGVYRVGIDQTKGLTTPKLTGVNNGNGKIVIDSQNGQWTI 1227			
DB	580	-----QIALNDK--SKGDVRVD--GONSLETTFNVYGT-----SGTGTL 615			
QY	1228	TGLSNTLANVTND-----KGSVYRTTEQNIILKDEDKTRAASIVDVISAGFNLOGNEA 1280			
DB	616	TLTNGTLNVEGGEVYLGVRPEAVGTNLIG--AAHGEAAMADAGFITNATKVEFGL---GEG 671			
QY	1281	VDFVSTYDTVNFADGNATTAKVTYDDTSKTSKVYDV-----NVDDT-----TIEVKDKKL 1331			
DB	672	V-FVFNH--TNNSDAGYQVMDLITGDD--KDGKVIHDAGHTVFNAGWTYSCKTL-VNDGLL 726			
QY	1332	GVKTTTLTS--TGTGANKFALSNOATGDALVKAQSDIVAH--LNTLSGDIQATKAGSOAUNS 1388			
DB	727	TIASHTADGVTCMGSEVVIANPCTLDLASTNSAGDYTLTNALKGDLGMLRVQLSSSKM 786			
QY	1389	AGYVDADGNKV-----IYDSTDNKYQAOKNDCTVDKTKVEAKD-----KLVAQAQTPDCT 1438			
DB	787	FGTHATGTGEFAGVLAQLKDS-----FTLERDNTAALTHAMQLQSDSENTTSVKVGEOSIGG 842			
QY	1439	LAQMNVKSVINKEQVNDANKKOGINEDNAFVKGLEKAASDN--KTKNAAVT-VGDL----- 1491			
DB	843	LA-MNGGTIIIFDTPATLAEGYISVDTLVVG--AGDYTWKGRNYQVNGTGDVLIDV 897			
QY	1492	-----NAVAQTPLT-----FAGDTGTAKKL-----GETLITKGGQTD-----TNK 1527			

```
Db 898 PKPNDPMANNPLTTLNLEHDDSHVGVQLVKAQFTVIGSGGSLTLRLDQDEVEADKTLH 957
QY 1528 LIDNNIGVAGTDGFTVKLAKD-----LTNLSVNA-GETKIDDKGVSVFSDSSGQAKANT 1581
Db 958 IAQNGTVAEGDYGFRLLTAPGNGLVYVGLKALNIHGGOKLT-----LAEHGGAYGAT 1011
QY 1582 PVLISANGLDLGGKVISNVKGKTKDTPAANVQOLNEVRNLLGLGNAGNADGNQVNIADI 1641
Db 1012 ADMSAK---IGG-----EGDLA-----INTVRQ-VLSLNGQNDYOGATYVQMGT 1052
QY 1642 KKDPSNGSSSNRTV-IAKGTVLGGKNDTEKLTATGQVGV-----VDKDGANAN 1689
Db 1053 RTDADGALNTRNLNSAIVDLNCTQTVEFTG--QMGSTVLKFGKALTVNKGGISQ 1110
QY 1690 GDLS-----NWVYKTKQDKSKKALLATYNAAGTNYLTN-NPAEAIIDRINEQGI-RFFHV 1742
Db 1111 GELTGGNLLNV---TGTLAIEGLNARYNA-----LTSISPAEVSNDLTQGLGRGNIA 1161
QY 1743 NDCNQEPVQGRN---GIDSSASGKHSVAIGFOAKADGEAAVAIGRQTQAGNQSIAIGDN 1799
Db 1162 NDG-----LTLKVTGTGLNLSISGKIVSA--TARTDVELD-----GDN 1199
QY 1800 AOATGQDSTAIGTGNVNVAGKHSAGIDPSTVKADNSYSVGNNOFTDATOT----- 1850
Db 1200 SRVFG--QFNIDTGSALSVEQKNGLDASVI-----NNGLLTISTERSWAMTHSIS 1248
QY 1851 ---DVEGVGNNTVTSNS-----VALGSNSAIS-AGTHAGTQAKKSDGTAG 1893
Db 1249 GSGDVTKIGTGLTLNDSAAVGGTTDIVGGEIATGSDSAINMASQHINIH--NSGVMSG 1306
QY 1894 TTTTACATGTVKGFACQTA 1912
Db 1307 NVTTAGDMNMPG--GGTA 1323
```

## RESULT 15

```
WAPA_BACSU STANDARD; PRT: 2334 AA.
AC Q07833;
DC 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE WALL-ASSOCIATED PROTEIN PRECURSOR.
GN WAPA OR NI7G.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
[1]
SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93302506; PubMed=8316082;
RA Foster S.J.;
RT "Molecular analysis of three major wall-associated proteins of
RT Bacillus subtilis 168: evidence for processing of the product of a
RT gene encoding a 258 kDa precursor two-domain ligand-binding
RT protein."
RL Mol. Microbiol. 8:299-310(1993).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=168 / BGSCIAL;
RX MEDLINE=95219088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapa loci."
RL Microbiology 141:337-343(1995).
CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC MOTILITY, SECRETION OR DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC INTO THE MEDIUM.
CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
```

```
CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
CC MOTIF REPEATED 31 TIMES.
CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
CC SIMILARITY TO THE REPEAT IN E.COLI RHS GROUP OF PROTEINS (RHS-A-D).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L05634; AAA22883.1; -
CC EMBL; D31856; BAA06656.1; -
CC EMBL; D29985; BAA06260.1; -
CC EMBL; D83026; BAA11683.1; -
CC EMBL; Z99124; CAB15959.1; -
CC PIR; S32920; S32920.
CC Subtilisin; HG10797; wapa.
CC Pfam; PF02018; CBD_6; 1.
CC Cell wall; Repeat; Signal.
CC SIGNAL 1 28
CC CHAIN 29 2334 OR 32 (POTENTIAL).
CC DOMAIN 504 869 WALL-ASSOCIATED PROTEIN.
CC REPEAT 504 605 3 X 101 AA APPROXIMATE TANDEM REPEATS.
CC REPEAT 636 736 1-1.
CC REPEAT 769 869 1-2.
CC DOMAIN 1021 2139 1-3.
CC 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
CC X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
CC 2-1.
CC REPEAT 1021 1040 2-2.
CC REPEAT 1042 1061 2-3.
CC REPEAT 1063 1082 2-4.
CC REPEAT 1083 1102 2-5.
CC REPEAT 1109 1128 2-6.
CC REPEAT 1129 1148 2-7.
CC REPEAT 1150 1169 2-8.
CC REPEAT 1174 1193 2-9.
CC REPEAT 1199 1218 2-10.
CC REPEAT 1219 1238 2-11.
CC REPEAT 1646 1665 2-12.
CC REPEAT 1667 1686 2-13.
CC REPEAT 1690 1709 2-14.
CC REPEAT 1711 1730 2-15.
CC REPEAT 1732 1751 2-16.
CC REPEAT 1753 1772 2-17.
CC REPEAT 1795 1814 2-18.
CC REPEAT 1820 1839 2-19.
CC REPEAT 1840 1859 2-20.
CC REPEAT 1861 1880 2-21.
CC REPEAT 1887 1906 2-22.
CC REPEAT 1908 1927 2-23.
CC REPEAT 1929 1948 2-24 (APPROXIMATE).
CC REPEAT 1969 1982 2-25.
CC REPEAT 1983 2002 2-26.
CC REPEAT 2008 2027 2-27.
CC REPEAT 2028 2047 2-28.
CC REPEAT 2051 2070 2-29.
CC REPEAT 2071 2090 2-30.
CC REPEAT 2093 2112 2-31.
CC REPEAT 2120 2139 2-31.
CC SEQUENCE 2334 AA; 258329 MW; B75138CCD278BAA3 CRC64;
```

```
Query Match 3.6%; Score 370.5; DB 1; Length 2334;
Best Local Similarity 20.1%; Pred No. 2.5e-06;
Matches 452; Conservative 242; Mismatches 771; Indels 787; Gaps 117;
QY 296 RPAVTPNTQALDPKFOATNNTK-AGPLSIGSNSIKRKRIINVAGVKNKT----DAVNYAQL 350
Db 349 KTGVDKTTCTGNAFMAFKFNLPQNMTVTATLKTVVAHSYCTKATGLWLDVNSNYD 408
QY 351 EAVVWAKERRITFCDDN--STDVKIG-----LDNTLTIK-----GGAETNALTDNNIGV 400
```

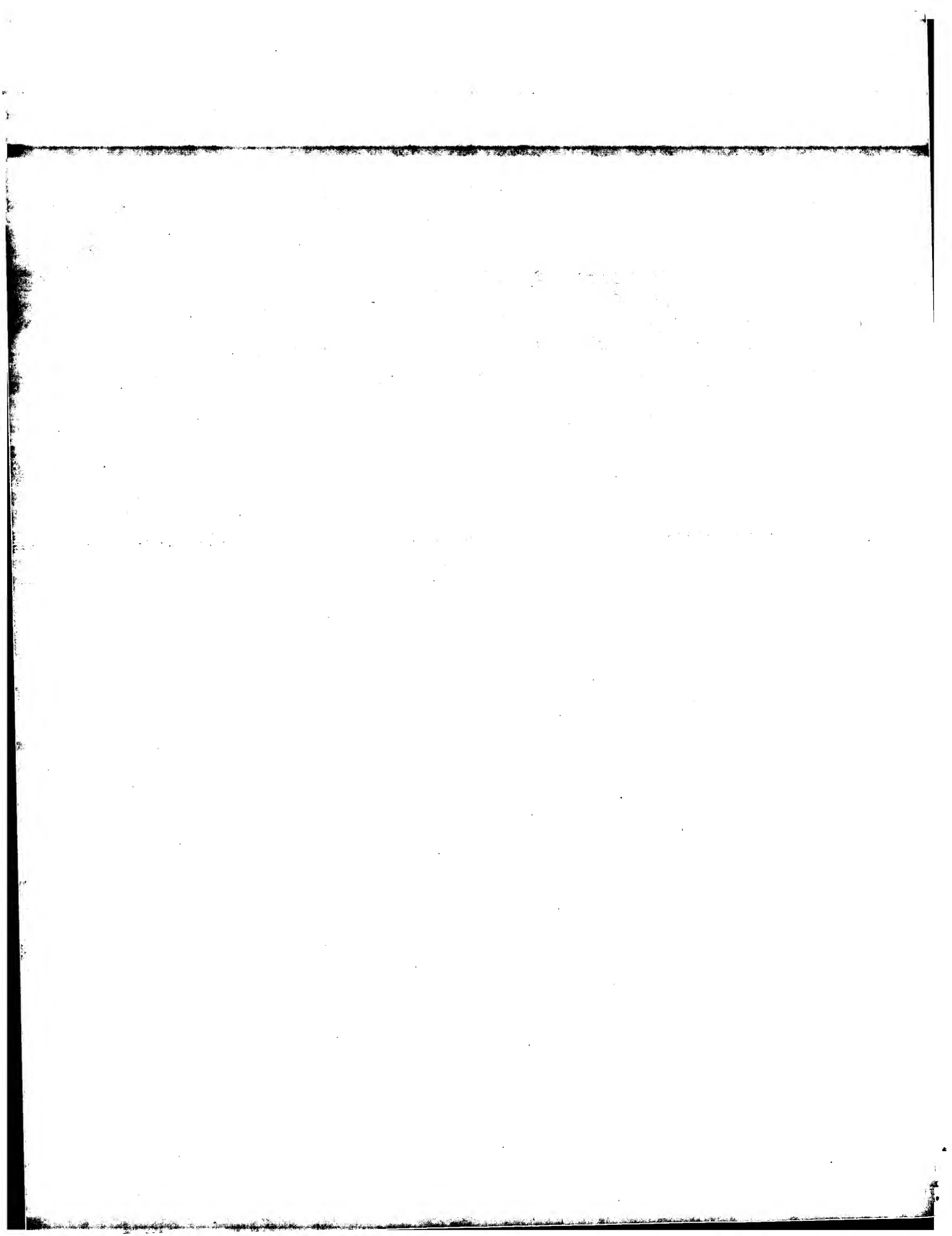
Db	1400	-----LAGNDWTKRQITFTTPANAGKAVVYNEVDHDKDCKGKRAWFDEVOLEKEV	1451
Qy	1197	-----QTKGLTTPKLTGVNNGKVIDSONCON-----TITGLSNT	1233
Db	1452	SSSNPVQSSSFTS---ATENMNVSGASVDEEGFDDVSLKAARTSASQAGSVTKQTVV	1508
Qy	1234	LANTVNDK-----GSVRTTEQNIKDEKTRAAASIVDVLSAGF---NLQNG	1278
Db	1509	LGOSANDKPVYLTGCMKASSVKFTDEKYSLOANVTYADGSTGLYNKAKPPSQEMNR	1568
Qy	1279	EAVDFVST-----YDTVPFADGNATTAKTVDYD-----TSKTSK	1312
Db	1569	AAVVPKTPINKVDISILFKSATSATGTVMFDDIRLIEGSLLT-KSTYDSNGNYVTKED	1627
Qy	1313	VYVDVND-----DFTIEVKDKKLGKVT-----STCTG-----ANKF	1348
Db	1628	LGAYATSDYDETKKTSSETDAK-GEKTTYTYDQADQLTNMTLSNCTSLHSHYDKEGNEV	1685
Qy	1349	ALSNQATGD-----ALVKASDIVAHLNLTLSGOIQTKAGSQAANSAGYVDADGN	1397
Db	1686	SKTIRAGADQTYKFEYDVGMKLVKTTDPLG--NVLASEYDA-----NSNLTKTISPNGN	1737
Qy	1398	KV--IYDSTD--NKYYQA--KNDGTVDKTKVAKDKLVAQAQTPDGTILAQMNVKSVINK	1450
Db	1738	EVLSYDGTDRVKSSYNGTEKTIIFYDKNG-----NETSVVVK	1776
Qy	1451	EQVNDANKKQGINEDNAPVKGLEKAAADNKTNAAVTVGDLNAAVQTPUTFAGDTGTTAK	1510
Db	1777	EQ--NTTKRTFEDKNKRLTDLTDRGS-----QT-WTYPSD-----SD	1811
Qy	1511	KLGETLTKGGQDTHKLTDDNIGVVGAGTGTFTVKLAKDLTLNLSVYNAGGTIKDDKGVSF	1570
Db	1812	KLKTFWIHGDKGTQFTYNTKIDQMIIE-----MKDSTSYSD-----	1850
Qy	1571	VDSGOAKANTPVLSANGLDLGKVISNVGKGTDDAANVOQLNEVRNLLGLGNAGNDN	1630
Db	1851	YDENGWQT--FITGNG--GGTSFS-----YDERNLVSLSHIGDKN	1887
Qy	1631	ADGNQVNIADIKDPNSSSSNRTVIK---AGTVL--GGKGNNDTEKLTAGGQVGVKDK	1685
Db	1888	--GGDILTESYEYDAN---GNRTIINSSASGKVQYEGKLNQLVKTHEDGTVIEVTD	1941
Qy	1686	GNAGDLNVWVKTKQDKSKALLATYNAAGQTNLYNNPAEADRINEOGI-----	1737
Db	1942	GFGN---RKTVTTIKDGSSKTVNASFINMQ-----LTKYNDESISYDKNGNR	1986
Qy	1738	---RFFHVD-----GNOEPVW---QGRNGIDSSASGK-----HSVAI	1769
Db	1987	TSDGRTYTWDAEDNLATVTKKEDKPPATYKDEKGNRIQKTVNGKVTNYFYDGDLSNV	2046
Qy	1770	GFOAKADGEAAVAIGRQTAQGNQSIAGDNAQ-----ATGD-QSIAIGT	1812
Db	2047	LYETADNNV-----TKSYTYGDSGOLLSTYENGKKFYFVHNAGHDIIAISDST	2095
Qy	1813	GNVAGKHSAGTGDSTVKA-----DNSYSYVNNNQFDTAT-----QTDVFE--	1853
Db	2096	GKTVAKYQDAMGNPTKTEASDEVKDNRYA-GYOYDEETGLLYLMARYEPNPGVFLS	2154
Qy	1854	---GYCNNITVTESVALGNS-----AISAG--THAGTQAKSDGTAGT	1894
Db	2155	LPPDGSGLDQNGYAYGNNPNVNNVDPDGHVWLVVYNAGFAAYDGYKAYKSG-----	2209
Qy	1895	TITTAGATGVTKFAGOTAVGVSVCASGAEIRIQNVAAGEVSATSTDVNGSOLYKATQS	1954
Db	2210	-----KMKGAWAAA-----SNFPGKIFK-----GASRAKFTKK	2241
Qy	1955	IANATNELDHRINON-----ENKANAGISSAMAMASMPQA---YIPGRSMVTGG	2000
Db	2242	AVKITGHTRHGLNQSLCRNGGRCVNLRAKINA-VRSPKKVVIKOPNGCATKYVGKATV---	2297
Qy	2001	IATHNGOGAVAGLSKLSNDQWVFKINGSAD	2032
Db	2298	VLNKRGKVIATAYGSSR-AKGSKHVFHTHGKN	2328

Db	409	NAKYTNW---TKPASKNIGKADVHKQWASVDVTAAYKSWNSGGA-----NYGFK	455
Qy	401	KEADNSGLKVKLAKTLNLTAEVNTTTLNATTVKVGSSSTTAELLSDSLTFPTQNTGSQ	460
Db	456	LHTNGNG-KEYWKKLISSANSANKPYIEVYTIPTKGNTPTIKAYHNGDS-----TGXF	507
Qy	461	STS-KTVYGVNGVK-FTNNAETTAAGTTRIT-----RDKIGF-AR	498
Db	508	DISWKKVEGARGYKWIYNGKEYQALISAGNVTSWSTGKKLWPTSAELASKRYKLHLDDGK	567
Qy	499	DG-----DVEKQAP--	508
Db	568	DGAELALDPSVPYKNSGSGYATSKNYWIGVSAIFDQEGEGAMSAKAPVPIPVNKAQAPSA	627
Qy	509	-----YLDKKQLKVGVA-----ITIDNG-----IDAGNKKISNLAKSSANDAVTI	550
Db	628	KYNNNGATGYFDLSWKAIVSGATGYKVQVFGKGFETLDELGN-QTSWTTGKKKI--WPTS	684
Qy	551	EQLKAAPTLN---AGAGISVTPTEISVDKSGNVTAPTY-----NIGVKTEL	596
Db	685	AETKAGYALHLKDGSGAELPINP-----GPTYKNAGGCGAKRNSFKILAY	731
Qy	597	NSDG-----TSDFSVKSGGTNNSLVTAEHLASYLN-----EVN	630
Db	732	NKDGEAISPAAATPALPIARPKNVTGLYTN---TKSSQTGYVNLWEKYQNAKGYKN	788
Qy	631	RTADSALQSTVKKEEDDDANAITVAKDTKN-----AGAVSILKLKGKNGLTV	679
Db	789	IYNGKEYQSFVGDAD-----HWTQKNIWPTSEEEKAGSYK-LHTDGGKGLA	837
Qy	680	AT-----KKDGTVFLQSQSLTIGKSTLN---NDGLIVKOTNEQIQV--	720
Db	838	LDPSPVNNANGYKCKNYSEFTLVAIDANGETIPTAPNPTFHEGAEFLGTEETWSIID	897
Qy	721	-----GANGIKFTN-----VNGSNPPTGIANT--ARITRDKI-----GFAGSDGAV--	759
Db	898	IPSGQLGATGNVIVNEEDLSIDRGPGLSRTYNSLSDSHLFGOGWYADAETSIVST	957
Qy	760	DTNKPYLDDQKLVGNVKTITNTGINAGG--KAITGLSPILPSIADOSSRNIELGNTIQDK	817
Db	958	DGGAMVIDEATTHRETFK-----KADGTYPQPTGYLELTETADQF-----ILKTK	1003
Qy	818	DKSNAASINDLNTGNLK-----NNNPNIDFVSYD-----IVDFANGNATTA	861
Db	1004	DOTNAY-----FNKGGKLOKVVDGHNAT--VYTYNDKNQLTAITD-ASGRKLTF	1051
Qy	862	T-----VTHDTANKTSKVYDV-----NVDDTTHLTGTD-DNKKLGKVTTKLAKTSA	908
Db	1052	TYDENGHVTSITGPKNKVTSYENDLLKKVTDGTDTGVTSYDSDSEGRLVKYQYSANSTEA	1111
Qy	909	-----NGWTAENFNVNSDEDALV---NAKDLAENLTLAKEIHTT	946
Db	1112	KPVTEYOYSGHRLKAEKNAKETTVYSYDADKTLTMTQPNRKGKVOYGYNEAGNPQVI	1171
Qy	947	KGADTALQTFVTKVVDENNADDA--NAITVGQKNANNQVNTLTKGELNLIKTDKNG	1004
Db	1172	DAAEGLKITWT--KYEGNNVVEDVDPNDVGTGKATESYQYD-----KDGNTSVK-DAYG	1224
Qy	1005	VTVFGINTTSLGKAGKSTLND-----GGLSIKNPTGSEQIOVGADGVKFAKVNNGNVWG	1058
Db	1225	TETYEYNNNDVTKMKDTEGNTVDIAYDGLDAVSET--DQSGKSSAAVYDKYGNQIOSS	1282
Qy	1059	AGIDGTRITRD---EIGFTGTNGSLDKSKPHLS-----KDGINAGGKKTNIQSGEIAQN	1111
Db	1283	KDLSASTNLKDGSEFAQKSGWNLASKDRRKISVIADKSGVLSGSKALEVLSOSTSAGT	1342
Qy	1112	SHDVTGGKIYDKLELENKISSTAKT---RQNSLHEFSVADEQGNPT--VSNPYSSYDT	1167
Db	1343	DHYSSATQVLEPNPTVTLTSGKTKDLSAKRAYFNIDLDKDKQKRIOWIHNEYS---	1399
Qy	1168	SKTSDVITFAGENG-----ITTKVN--KGVYRVGID-----	1196

Search completed: September 13, 2001, 12:52:48  
Job time: 798 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:42:30 ; Search time 112.99 Seconds  
(without alignments)  
2396.924 Million cell updates/sec

Title: US-09-361-619-7  
Perfect score: 10303  
Sequence: 1 MNHIYKVFKNKATGTFMVA.....NGSADTQGHVCAAVGACGPHF 2047

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_16:\*  
1: sp-archaea:\*  
2: sp-bacteria:\*  
3: sp-fungi:\*  
4: sp-human:\*  
5: sp-invertebrate:\*  
6: sp-mammal:\*  
7: sp-mhc:\*  
8: sp-organelle:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-unclassified:\*  
13: sp-vertebrate:\*  
14: sp-virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1143	11.1	2059	Q9PD50	Q9pd50 xylella fas
2	1123	10.9	2353	P71401	P71401 haemophilus
3	1046.5	10.2	2712	Q9F3X5	Q9f3x5 pasteurella
4	778	7.6	1190	Q9PC04	Q9pc04 xylella fas
5	739	7.2	1299	Q9F3X6	Q9f3x6 pasteurella
6	678.5	6.6	1098	Q48152	Q48152 haemophilus
7	661	6.4	1107	Q9F2D8	Q9f2d8 salmonella
8	627	6.1	3705	Q9F2D8	Q9f2d8 salmonella
9	582	5.6	2340	Q9F2D9	Q9f2d9 yersinia pe
10	570.5	5.5	4919	Q9ZHL0	Q9zhl0 haemophilus
11	567.5	5.5	2021	Q52657	Q52657 rickettsia
12	555	5.4	2106	Q9XC47	Q9xc47 rickettsia
13	548	5.3	2586	Q9VTK8	Q9vtk8 drsophila
14	548	5.0	3029	Q53582	Q53582 synchocyst
15	514.5	5.0	2035	Q9XCJ4	Q9xcj4 salmonella
16	514	5.0	1604	Q9KK99	Q9kk99 rickettsia
17	512.5	5.0	1557	Q9RNI2	Q9rni2 haemophilus
18	509.5	4.9	5627	Q9I120	Q9i120 pseudomonas
19	506	4.9	4152	Q9ZHL3	Q9zhl3 haemophilus

20	502	4.9	2747	2	Q9L800	Q9l800 aeromonas s
21	499	4.8	1758	2	Q9JMS5	Q9jms5 escherichia
22	494.5	4.8	2232	5	P91365	P91365 caenorhabdi
23	491.5	4.8	13288	6	O18758	O18758 sus scrofa
24	485	4.7	2478	2	Q9LCH2	Q9lch2 staphylococ
25	483	4.7	2478	2	Q9RL69	Q9rl69 staphylococ
26	479	4.6	2468	2	Q9I2M3	Q9i2m3 pseudomonas
27	478.5	4.6	2349	2	P94750	P94750 escherichia
28	478.5	4.6	2383	2	P76347	P76347 escherichia
29	477.5	4.6	2514	2	Q9JY30	Q9jy30 neisseria m
30	469	4.6	1963	2	Q9XCQ3	Q9xcq3 salmonella
31	468	4.5	1615	2	Q9KKA4	Q9kka4 rickettsia
32	467	4.5	1255	2	Q9FDA0	Q9fda0 xanthomonas
33	466	4.5	1975	2	Q9K057	Q9k057 neisseria m
34	465.5	4.5	1616	2	Q9KKB8	Q9kkb8 rickettsia
35	462	4.5	2893	2	O25063	O25063 helicobacte
36	456.5	4.4	1029	2	O52708	O52708 rickettsia
37	456	4.4	3381	2	Q9KX33	Q9kx33 streptococ
38	455	4.4	1643	2	Q9F0P7	Q9f0p7 rickettsia
39	455	4.4	2147	2	Q9L950	Q9l950 pseudomonas
40	453	4.4	1615	2	Q9F0P9	Q9f0p9 rickettsia
41	452.5	4.4	1619	2	Q9KKB7	Q9kkb7 rickettsia
42	452	4.4	1328	2	Q9LAX0	Q9lax0 xanthomonas
43	452	4.4	1643	2	Q9F0P6	Q9f0p6 rickettsia
44	452	4.4	2055	2	O85472	O85472 abiotrophia
45	452	4.4	2703	2	Q9K0T0	Q9k0t0 neisseria m

## ALIGNMENTS

RESULT	1
Q9PD50	
ID	Q9PD50 PRELIMINARY; PRT; 2059 AA.
AC	Q9PD50:
DT	01-OCT-2000 (TREMBLrel. 15, Created)
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE	SURFACE PROTEIN.
GN	XF1529.
OS	Xylella fastidiosa.
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC	Xylella.
OX	NCBI_TaxID=2371;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	SPRAIN=9A5C;
RX	MEDLINE=20365717; PubMed=10910347;
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA	Faccinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA	Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA	Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA	Mench C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA	Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA	da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
RA	de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA	Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,



RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

RT "The genome sequence of the plant pathogen Xylella fastidiosa.";

RL Nature 406:151-159(2000).

DR EMBL; AE003982; AAF84338.1; --

SQ SEQUENCE 2059 AA; 204035 MW; EBA6A71B5DD24E10 CRC64;

Query Match	11.1%	Score 1143;	DB 2;	Length 2059;
Best Local Similarity	24.0%	Pred. No. 9.6e-40;		
Matches 547;	Conservative 342;	Mismatches 766;	Indels 626;	Gaps 104;
Qy 86 SGT A--KADGDRAIAIGENANAGQQAIAIGSSNKT--VNGSSLDKIGTGOESTAIIG 141				
Db 86 SGTAAEQGASRNLTGGSLYVNSGQGVNDVNLKNTYSIRMGSVTMTVA--GNATAIG 144				
Qy 142 GDVKASGDASIAIGSDLLHLDHCHNPKHPKGLTLNDLNGHVLKEIRSSKDNVDKYRR 201				
Db 145 SAQSSAADAL-----KASL-----A 159				
Qy 202 TTASGHASTAVGAMSYAQGHFNAFGRTRATAKSAVLAAGLAATAEGOSTIAIGDATSS 261				
Db 160 TKASGARAIAGAKASAGDGVTVALGSGATA-----GTGASSIAIGLNASAV 206				
Qy 262 SLCAIALGAGTAAQL--QGSIALGOGSVVTVQSDNNRSPAYTPNTQALDPKFOATNNTKAGP 320				
Db 207 N-GAVAVGGGALVTVDPGAVAGLNSVA--STGKLGSGVDPKTKTSTSDASAANKSTLAA 263				
Qy 321 LSTG---SNSIK-RKTIYVAGVYKNTDVAVVAOLEAVKWKAKERRITPOGDNDSTDKIG 376				
Db 264 VSGDVSSTNLKTRQLSGLAAGTSNTDAVNAQLKVVDEIASR-----306				
Qy 377 LDNLTITKGAETNALTDDNNGVGVKADNSGLKVLAKTLNLTETVNTTLNATTVKVG 436				
Db 307 -GMNLTASGANGNAPGSSV-----DLKNTDKNLTITKAIG 342				
Qy 437 SSSSTTAEALLSDSLTFQPTQNTSQSTSKTVYGVNGVKFTNNAP--TTAAIGTTRITRDK 493				
Db 343 S-----NDVQFNLNKDKVTVTLAVGDALLNTDG 370				
Qy 494 IGFARDGVDDEKQAPYLDKKQLKVGSAIT-----IDNGIDAGNKKISNLAKGS---SAN 545				
Db 371 IALGTD-----VSLSTTGIAITDGPNAVATASGIDAGSKVISHVAGAVSETST 417				
Qy 546 DAVTEOLKAAKPTLNAGAGTSVPTPEISVDAKSG-----NVTAPYINIG-VKTEL 596				
Db 418 DAVNGSQLNAVQ--VQASQPTFTGNEGAVKRSGLQGSVWISGESSTAGTYSGGNLKASVVD 475				
Qy 597 NSDG-----TSOKFS--VKGSGTNNSLVTAHLSYLNVEVNETADSAQSFTVKEEDD 647				
Db 476 EAAGRIHLQALDSPKFGNVVINGGKISGVTAG-----TEETDAVNFSQLKSISTAV--- 527				
Qy 648 DDANAITVAKDTTKNAGAVSILKLGKNG-LTVATKKDGT-VTFGLSQD---SGLITGKS 702				
Db 528 DQWTLTASGNGSKVASGTVDLKNTDGNLTISKSGSDNDVFNLSKDFKVDGMTSGTT 587				
Qy 703 TLNNDGLTVKDTNEQIQVANGIKFTN-----VN 731				
Db 588 VVNDGVKV---GSDVALGTGLTITDGPNAVATASGIDAGSKVISHVAGVVSSTDAVN 644				
Qy 732 GSNPGTGANTARITRDK-IGFAGSDGAVDTPKPYLDQDKLOVGNKVTINTGINAGKAI 790				
Db 645 GSO-----LNAVQVQASQVPTFTGEGAV---KRSLGVSIVISG--ESSTAGMYSGG--- 691				
Qy 791 TGLSPPLPDIADQSSRNIEL-----GNTI-----QDKKSNA--SI 825				
Db 692 ----NLKSVDEAAGRIHLQALDSPKFGNVVINGGKISGVTAGTEETDAVNFSQLKSI 746				
Qy 826 NDLNTGNFNLKNNPIDFVSFYDVIDFANGNATTATVTHDTANKTSKVYV----DNVVD 881				
Db 747 STAVDQGWTLTASGNSKVASGTVDLKN---TDGNLTISKSGSDNDVFNLSKDFKVD 803				
Qy 882 DTTIHLTGT---DDNKKLG---VKTTLKNTKSANGNTATNFN-----VNS 920				

Db 804 GMT---SGTTVVNDGVKVGSDVALGTGTLTITDGPNAVATASGIDAGSKVISHVAGVVSE 860				
Qy 921 SBDAL---VNAKDIAEN-----LNTLAKET-----HTTKGT-----A 950				
Db 861 TSTDVAVNGSQLNAVQVQASQPTFTGNEGAVKRSGLQGSVWISGESSTAGTYSGGNLKSVV 920				
Qy 951 DTALQTFTVKKVDEN-----NNADDANAITVGOK-----NANNOVNTLT 989				
Db 921 DEAGRIHLQALDSPKFGNVVINGGKISGVTAGTEETDAVNFSQLKSISTAVDQGWTLT 980				
Qy 990 LKGENGLINKTKNGIVTPIGINTTSLKAGKSTLNDGGLSINKPTGSEOI-----OVGA 1043				
Db 981 ASGANG--SKVASGGTVDL-----KNT--DGNLTISKSGSDNDVFNLSKDFK 1025				
Qy 1044 DGKFEAK--VNNNGV-VGAGID-GTTRITRDEIGFTGTNGSLDKSKPHLSKDGINAGKK 1099				
Db 1026 DGMTSGTTVVNDGVKVGSDVALGTT-----GLTIANG-----PAVTASGIDAGSKV 1072				
Qy 1100 ITNIQSGEIAQNSHDVATGGKIYDLKTELENKISSSTAKTAQNSLHEFSVADEQGNFTVS 1159				
Db 1073 ISHVAAGAVSETSTDAVNGSQLNAVQVQASQPTFTG-----NEGAVKRSGLQSVWIS 1125				
Qy 1160 NPYSSTYDTSKSDVITPAGENGITTKVNGVVRVGDQTKGLTTPKL--TVGNNNKGIV 1217				
Db 1126 -----GESSTAGTISGGN-LKSVVDEAAGRIHLQLA--DSPKFGNVVINGGK--- 1170				
Qy 1218 IDSQNGQNTITGLSNTLANVTNKGSVRTTEQNI IKDEDKTRAASITVDLSAGENLQGN 1277				
Db 1171 -----ISGVF-----ACETEEDAV---NFSQKSISTAVDQGWTLTAS 1205				
Qy 1278 CEAVDFVSTYDVFNF--ADGNATTAKVYDDTSTKSVVYDVNVVD-----DTTIEVK 1327				
Db 1206 GANGSKVASGGTVDLKNTDGNLTISK-----SGSDNDVFNLSKDFKVDGMTSGTTVVNN 1360				
Qy 1328 DKILGVKTTTLTSTGTGANKFALSNQATGDLVKASDIIVAHNLTLSDGIQTAKGASOANN 1387				
Db 1261 D---GVKVGSDVALGTGTLTIANGPAVTASGIDAGSKVISIV-----AGAVSETS 1308				
Qy 1388 SAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKREAVAKDLVAQAQTPDGTLAQMNYKSV 1447				
Db 1309 TDVAVNGSQLNAVQVQASQPTFTG-NEGAVKRS--LQGSVVISGESSTAGTYSGGNLKSV 1365				
Qy 1448 INKEQ-----VNDANKKQGT---NEDNAF-----VKGLEKASDNKTK 1482				
Db 1366 VDEAAGTIHLQALDSPKFGNVVINGGKISGVTAGTEETDAVNFSQLKSISTAVDQGWTL 1425				
Qy 1483 NAAVTVG-----DL-NAVAQTPLTAFAGTGTGTTAKKLGTLTIK-----G 1520				
Db 1426 TASGANGSKVASGGTVDLKNTDGNLTISKSGSDNDVFNLSKDFKESITVGTOLDKDG 1485				
Qy 1521 GQTDNKLTDNNIGVAVAGTDGP---TVKLAKDL'NLNSVNAAGTKIDDKGVSVFVDSGQ 1576				
Db 1486 VKVSSNVLLDSNELVITSHSSTSSVKTLANGBSVVNRTVVNGDGVNIDB---VVVYNDLGL 1543				
Qy 1577 AKANTEVLSANGLDLGGKVISVNGKGTGTDDAANVQOLNEVYNLNLGLG- NAGNDNADGNQ 1635				
Db 1544 SIVGGASLTLSGINAGSHKITVNTAGTEDDAVNFSQLKSSEAVDKGWTLTASGANGSK 1603				
Qy 1636 V---NIADIKK-DPN---SGSSSNRTVI-----KAGTVLGGKGNNDTEKLTATGGIQVG 1681				
Db 1604 VSGGTVDLKNTDGNLAIKSGSDNDVFNLSKDFKVDVETAG-----NTVNTDGVKVG 1658				
Qy 1682 VDKDGNANG-----DLSNVVWVKQDKGSKALLATYNAAGQTNLTNPA---EATDRINE 1734				
Db 1659 SDVSLCAGMLFTANGPSVTSAGFNAGDK---VISHVAVGMAOTDAVNVSQLKQAVQSVTV 1715				
Qy 1735 QGIRFFHVNDGNQEPVVGGRNGIDSSASGKHSVAIGFQAKADGEAAVAIGROTQA--GNOS 1793				
Db 1716 KATRYSTNDGG-----TOGNYDGDGATGSKALTAAGVGTQASGEGAAVSGSAAASGKGS 1771				
Qy 1794 IATGDNAQATCDOSIAIGTG-----NVVAKHSG-----AIGDPSTVAKADNSV 1836				
Db 1772 TAIGRNAIASADGSVALDGAKDKGGRGAESYTGKYSQVQNTVTGTVSVGDAKGETRSIS 1831				

```
QY 1837 SVGNNOFTDAT-----QT-----DVFGV--GNNTVTVES-----NS 1866
| : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1832 NVADAKEAMDANRLQRLDVAQKSNLQTDHRRHEINNIEDVFKITKGDSSASSVKGMGVNA 1891
| : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1867 VALGNSAISAGTHAGTOAKKSDGTACTTTTACATGATVKGFGAGTAVGAVSVGASGAPRR 1926
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1892 MATGTNAVS--GTESVALGKNTWNSADNAVAG--NGSVADRA-----NSVSVGGGSRQ 1944
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 1927 IQNVAAGEVSATSDAVNGSOLYKATQSIANATNELDHRHONENKANAGISSAMAMASM 1986
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1945 VTNVAAG---TADTDVNVNSQL---NQGLITAKQYTDGMVGNLRRETSGGVAIAIATANL 1998
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 1987 POAYIPGRSVMTGGIATHNGGAVAVGLSKLSDNCQWVFKINGSADTQGHVGAAGACPH 2046
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1999 POAYVQGRGTSVGVSSYQGSIAVGSVAVSESGHWFVFKSGSANTRSHVGVGAGVGYQ 2058
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 2047 F 2047
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 2059 W 2059

RESULT 2
P71401 PRELIMINARY; PRT; 2353 AA.
ID P71401
AC P71401
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE HSF PROTEIN.
GN HSF.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C54;
RX MEDLINE=97047989; PubMed=8892830;
RA Geme J.W., Cutter D. III, Barenhamp S.J.;
RT "Characterization of the genetic locus encoding Haemophilus influenzae
RT type b surface fibrils.";
RL J. Bacteriol. 178:6281-6287(1996).
DR EMBL; U41852; AAC44560.1; -.
SQ SEQUENCE 2353 AA; 243844 MW; 53E491B51A897B77 CRC64;

Query Match 10.9%; Score 1123; DB 2; Length 2353;
Best Local Similarity 23.08; Pred. No. 7.9e-39;
Matches 607; Conservative 318; Mismatches 838; Indels 874; Gaps 119;

QY 1 MNHIKYIFNKATGTFMVAEYAKSHSTGGSCATGQGVSVCTLSFARIAALAV----- 54
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 MNKIFNVIMVMTQTVVWVSELTRTHK--RASATVETAVLATLLEFATVQANATDEDEEL 58
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 55 -LVIGATLSSSAYAAQKDXTHIAIGEQNPPRS-----GTAKADGDRATIAIGNAN 104
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 59 DPVVRTAPVLFSHDSKEGT-----GEKEVTENSNGWYFDNKGVLKA-GAITLKAGDNLK 112
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 105 AOGGQAIAGSSNKTAVNGS---SLDKIGTGDAT--GQESIAIGDVKASGD-ASIAIGSD 157
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 113 IK-----QNTDSTWASSFTSLKKDLDTLSVATEKLISFG-----ANGDKVDITSAN 161
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 158 DLHLLDGHGPKHPKG--TLINDLINGHVLKEIRSSKDNMDVKYRRT----- 202
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 162 GLKLAKTGNVNLGLDSTLPDVAVTWGVLS--SFTPNVDVETRAATVKDVLNAGWNI 220
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 203 ----TASGHASTAVGMSYAGHPFSNAGFTRATAKSAYSIAVGLAAFGOSTIAIGSDA 258
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 221 KGARTAGGNVESVLDVSAYNVFEIT--GDKNT-----LDVVLTAKEGKTEVKEFTP 272
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 259 TSSSLGAIAGAGTGAQLOGSIALGQGSVVTSQSDNNSRPAYTPNQALDPKFOATNTKA 318
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

```
Db 273 TS-----VIKEKDGKLFPGCKENNDTNKVTSTNT-----ATDNTDE 306
QY 319 GPLSIGSNSIKRRIINYGAGVKNKTDVAVNAQLEAVVWAKERRITTFQDDNST----- 371
| : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 307 GNLVTAKAVIDAVNKAQRVKTITTTANQONGDEATV--ASGTNVTFESGDTGTASVTKDT 364
| : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 372 -----DVKIGLDNT--LTIKGG----- 386
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 365 NNGITVKYDAKVGDLKFDSDKKIVADTTALTVTGKVAIEAKEDDDKKLVNAGDLVTA 424
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 387 -----AETNALDDNI--GVVKEAD-----NSGLKVK-----LAKTLNN 418
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 425 LGNLSWAKAKADTDGALLEGISKQOEKAGETVTFKACKNLKVKQDGFANFYISQDALTG 484
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 419 LTEVNTTLNATTVKVGSSSTTAELLSLSLTFTQPNQTSQSTSKTV--YGVNGVKFTNN 477
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 485 LTSI---TLGGTTN---GGNDAKTV--INKDLGTTTPAGNGGTTGTNTISVTKDGKICAGNK 537
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 478 AETTAAG-----TTRTRDKIGFARGD--DYDEKQAPYLDKKQLKVGCSV 520
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 538 AITNVASGLRAYDDANFDVLNNSATDLNR--HVEDAYKGLLNLEKNA---RKQPLVTDST 593
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 521 AITIDN-----GIDAGNKKISNLAAGSSANDAVTIEQLKAARKPTLNAGAGISVTPTE 572
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 594 AATVGDRLKLGWVSTKNGTKEEN-----QVQADEVLFTGAGAA----- 634
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 573 ISVDAKSGNVTAPTYNIGVKTTE---LNSDGTSDKFSVSGSGTNNSLVTAHSLASYLNE 628
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 635 -TVTSKSEN--GKHITVSVVAETKADCGLEKDGDTIKLVQNDQNTDNLVTGN----- 684
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 629 VNRTADSALQSFVKE--EDDDANAITVAKOTTNAGAVSILKLGKNGLVAVTKKQDQTV 687
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 685 -NGTAVTKGGFETVKTGATDADRGKTV--KDATANDADKKVATVK-----DVATAINSAA 737
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 688 TF-----GLSODSGLTIGKSTLNDGLTVKDT---NEQIQVGANGIKFT----- 728
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 738 TFVKTENLTTSIDEDNPTDNGK---DDALKAGDTLTFKAGKNLKVKRDKGNITFDLAKN 793
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 729 -----NVNGSNPGTGIANTARI-----TRDKIGFAGSDGAVDTNKPILDODKLQ 772
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 794 LEVKTAKVSDPLTIGGNTPTGTTATPKVNITSTADGLNFA----- 834
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 773 VGNVKIINTGINAGGKAI--TGLSPTL--PSIADQSSRIELGNTIQDKSKNSAASINDI 828
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 835 -----KETADASGSKNYLKGAIATLTTEPSAGAKSS--HVDL--NVDATKKSRAASIEDV 885
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 829 LNTGFNLKNNNPIDFVSTYDIVDFANGNATATVTHDTANKTSKVYVDVNVDDTTIHLT 888
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 886 LRAGNIOGNNNDYVATYDTVNFNTDDSTGTTTIVT-----VTQKAD 927
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 889 GTDDNKKLGVTTKLTKNTSANGTAT-----NFNVNSSDED-----ALVNARDIAEN 935
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 928 KGKADVIGAKTSVK--DHNGKLFTGDKLDKDNANGATVSEDDGDKDGTGLVAKTVIDA 985
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 936 LNTLAKEIHHTKGTADALQFTTVKKYVDENNADNANAITVQO-----KNANNQVNTFLK 991
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 986 VNKSGWRVTGEGATAETGATAV-----NAGNAETVTSCTSVNFKNGN--ATTATVS 1034
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 992 GENG--LAIKTDKNGTVFINGINTTSLKAG-----KSTLNDGGLSIKNTGSEIQI-- 1040
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1035 KDNGININKYD-----VNVGDGLKIGDDKKIVADTTTUTVTGKVSVPAGANSVNN 1086
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 1041 ---VGADGVKPAKVN----- 1053
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1087 KKLVAEGLATA--LNNLSWAKADYADGESEGETDQEVKAGDKVTFKAGKNLKVQKQSEK 1145
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 1054 -----NGVVGAGIDGT-----TRITRDEIGFTGTNGS---LDKSKPH---LS 1089
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1146 DFTYSLODTLTGLTSITLGGTANGRNDGTVINKDGLTITLANGAAAGTADSNNGNTISVT 1205
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 1090 KDGINAGGKKTNTQSG---EIAONSHDVAVTGGKIYDLKTELENKI-----SSTAK 1137
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1206 KDGISAGNKEITNVKSALKTYKDTQNTADE--TQDKEFHAAYKNAVEFEVFGKNGATVSAT 1264
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

QY 1138 TAQNSLH-----EFSVADEQGNFTVSNPYSSDYTSK----- 1169  
Db 1265 TDNNKHVITIDVAEAKVGDGLEDKTDGKIKLKVDNTDGNLLTVDATKGASVAKGEFNA 1324  
QY 1170 -TSDVITFAGENG-----ITTKVNGVVRVGDIDQTKGLT----- 1202  
Db 1325 VTTDATTAQGTNANERGVVVGKSGNGATATETDKKKVATGVGVAKAINDAATFKVXEND 1384  
QY 1203 -----TPKLT 1207  
Db 1385 SATIDSPDTCANDALKAGDGLTLKAGKNLKVYKRDGNITFALANDLSVKSATVSDKLS 1444  
QY 1208 VGNNGK-----GIVDSQNGMT-----ITGLSNTLANVNDKGSVRTTQGNLIK 1254  
Db 1445 LGTNGKNVITSDTKGLNFAKDSKTDGDDANIHLNGIASTLTDTLNSGAT-TNLGGNIGIT 1503  
QY 1255 DEKTRAASIVDLSAGFNLOG-----NGEAVDFVSTYDTVNFADGNATKATVYD- 1305  
Db 1504 DNEKKRAASVKDVLNAGWVGVKPSANNOVENIDFVATYDTVDVFGDKRDTTSVTYES 1563  
QY 1306 -DTSKTSKVYDVNVDDTTEIVKDKLGVKTTLTSTG-----TGAN-KFALSNQAT----- 1355  
Db 1564 KNGKRTVE-----KIGAKTSVIKDHNGKLTGKELKDANNNGVTVTET 1607  
QY 1356 -----GDALVKASDIVAHLNLTLSGDIOTAKGASQANNSAGYVD-----ADGNKVLYDS 1403  
Db 1608 DGKDEGNGLVTAKAVIDAVNKAQWRYKTT-GANGQNDPATVAGSTNVTFFADNGTGTAEV 1666  
QY 1404 TDNKYYOAKNDG-TVDKTEKAVK-----DKLVAQA-----QTPDGTPLAQMNVKSVINKBO 1452  
Db 1567 T-----KANDGSTTVKYNVKVADGLKLDGDKIVADTVTLTVADGKVTAPN----- 1711  
QY 1453 VNDANKKOGINEDNAFYKLEKAASDNKTKNAATVYVDLNAVATPLTATGDTGTTAKKL 1512  
Db 1712 -NGDCKK-----FVDSASGLADALNKLKSWTA-TAGKEGTGEVDPAISAGQ-----EVKA 1757  
QY 1513 GETLTIKGGQTDNKLTDNNIGVGVAGDGTVKLAKDLTLNLSV-----NAG-----GTKI 1563  
Db 1758 GKVITFRAG-----DNLKIKQSKGDTYSILKELKDLTSVEFKDANGGTGSESTKI 1808  
QY 1564 DKGVSVFVSSGOA-----KANTPVLNANGLDLGGKVISNVGKTK-----DTDAA 1609  
Db 1809 TKDGLTITPANGAGAAGANTANTISVTKDISAGNKAVTVVSGLKFKFGDGHFLANGTVA 1868  
QY 1610 NVQOOL-NEVENLLGLGNAGDN-----ADGNQVNIAD-----IKKDPNSGS----- 1649  
Db 1869 DFEKHYNAYKDLTNLDEKAGDNNPTVADNTAATVGDRLGLGWVISADKTTGEPNQEYNA 1928  
QY 1650 ---SSNRTVIKAGT-----VLGG-----KGNNDTEKLTGCG-----LOVG 1681  
Db 1929 QVRNANEVFKRSGNGINVSGLTNGTRVITPELAKGEVKSNEFTVKNADGSETNLVKVG 1988  
QY 1682 -----VDKDNANGDLSNVVKTOK-----DGSKKALLATVYNAAGOTNYLTNN 1724  
Db 1989 DMYSKEDIDPATSKPMTG---KTEKYKVENGVKVSANGSKTEVTLNKGSG---YVTGN 2042  
QY 1725 PAEADIRNEOQIRFEHFHVDNGQEPVVOGRNGIDSSAGKHSVAIGFOAK----- 1774  
Db 2043 ---QVADATAKSGFEL-----GLADAAEAEKAFESAERKDKLSDKAETV 2084  
QY 1775 -ADGEAAVAIGROTQAGNOSIAIGNAQTGDO-----SIAIGTGNVAVGKHSAGIG 1825  
Db 2085 NAHDKYRFANGLNTKY---SAATVSESTDANGDKVTTTFVKTDVPLTQIY---NTDANG 2138  
QY 1826 DPSTVAKDANSYSGVNNNOFTDATQDVFQVGNNTVTSNVSVALGNSAISAGTHAGTQA 1885  
Db 2139 NKIVKADCKW-----VELNADGTASKEVTLGNVANGKK-----VV 2176  
QY 1886 KKSDDGTAGTGTATGTVKGFAGQATVAGVSVG-----ASSAERRIONVAG 1933  
Db 2177 KVTENGADKWKYTNADGAADKTKGEVSDNDKVTDEKHVVRDPNNQSGKGVVIONVANG 2236

QY 1934 EYSATSTDAVNGSOLY---KATQSIANATNELDHRHONENKANAGISSAMAMASMPQAY 1990  
Db 2237 ETSNTSTDAINGSOLYAVAKGVTNLAGQVNNLEGVKNVKGKADAGTASALAASOLPQAT 2296  
QY 1991 IPGRSMVYTGIIATHNGQAGAVGLSKLSQDSQWVFKINGSADTOGHVGAAGVAGGFHF 2047  
Db 2297 MPCKSMVAITAGSSYQONGLAIGVSRISDNGKVIIRLSCTTNSQGRKTGVAAGVGTQW 2353  
RESULT 3  
Q9F3X5 PRELIMINARY; PRT: 2712 AA.  
AC Q9F3X5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE MAPB PROTEIN.  
GN MAPB.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Pasteurella.  
OX NCBI\_TaxID=747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PM70;  
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;  
RT "Evolutionary origins of the autotransporter proteins."  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ277636; CAC14203.1;  
SQ SEQUENCE 2712 AA; 276154 MW; 3F5579D6F32FFA3D CRC64;  
Query Match 10.2%; Score 1046.5; DB 2; Length 2712;  
Best Local Similarity 22.9%; Pred. No. 1.5e-35;  
Matches 591; Conservative 348; Mismatches 917; Indels 725; Gaps 119;  
QY 7 VIFNKATCTFEMAVAEYAKSHSTGGGSCATQGVSVCTLSFARIAALAVLVIGATLSGSAY 66  
Db 309 VYDGASSVYVLPKDGESSVAIGDKAVSRGE-----ASIAIGK-----NAI 349  
QY 67 AOKKDTK-----HTAIGEQNPPRRSGTAKA-----DGDRAIATIGENANAAGGQAI 111  
Db 350 TSNDAKQSIQAKNNIALGTNAQAIASDNSIALGNARTNKDSDSIALCDSAEYKAAHSL 409  
QY 112 ATGSSNKTVNGSSLBKIG--TDTAQESTAIAGDVKASGDASIAIGSDDLHLDDO----- 164  
Db 410 AVGTTTSKALAEAL-ALGKLAETKTSVAMGNTSKADGNSVAVGNTSQTLSQNTIAIG 468  
QY 165 HGNPKHPKGLTLNDLINGHAVLKEIRSSKDNVDKYRRTTASG-----HAST-AV 212  
Db 469 SSAIANPERTISIGLNAGKQGEADATGTHSQINIGENSGEGVIGQLNIGIGHAGTNNV 528  
QY 213 GAMSVAQGHFSNAGFTRATAKSAISLAVGLAA-----TAEGQSTI-----AI 254  
Db 529 GKHNIALGTIYAGTNLKNSEETSGANVSIGHEANKYDOLTAQVOKSTVVGQAOTKAASRSTAL 588  
QY 255 GSDATSSSLGATAGTATRAQLOGSIALQGGSVVTVQSDNNSRPAYTPNTQALDPKFOATN 314  
Db 589 GAETALGLDAVAVGITSKAEGDKSVAIGANS--TADSNVALGATSRVAKEGSGYLIG 646  
QY 315 NTKAGPLSLIC-----SNSLKRIINVGAGVKNKTDVNVVAQLEAV----- 353  
Db 647 KQSSLVVSVGHQGAADQHLRLRVNVADGVDEQDAATVAQLKKVTEKYTSDLQAOFNAL 706  
QY 354 -----VKW-----AKERRITFOGD-----DNSTD-VKIGLDFLTIKGG 386  
Db 707 SHAPTASEIKYDTPVPSPTTGAENKITLAKTRISNVAPAEPLDVAINGVNHVITKNK 766  
QY 387 AETNALTDDNIGVVK-EADNSGLKVKL-----AKTLNNLTVEVNTTLNATTTVKYG 436  
Db 767 AHYFSVNDGTGINPVGPHNDNDGATAKLMAIGQNAAKAERSVAIGNNTVNGEGSIGLG 826  
QY 437 SSSSTTAEALLSDSLTFTQPTQNTGSQSTSKTVYGVN-GVKFTNNAETTAAGTTTITRDKIG 495

Db 827 TYVKSABE-LDDGVPRESETVTKPSKNVYKGAIGACTTTDGNNSIAIGSLAATSDK-- 883  
QY 496 FARDGDVDEKOAPYLDKKOLKQVSGVAITIDNGIDAGNKKISLNKAGSSANDAVTIEOLKA 555  
Db 884 -----NPGANVDRAIAIGYNAVSAFAKANAIGRAVANSVKG 920  
QY 556 AKPTLNAGAGISVPTPEISVDKSGNVTAPTNYGVKT----- 593  
Db 921 NAFGSQALSG-AESSTAIGTESSEGGNA--YALGTSHAKGLNSIAFGTNOVYSGQNSG 977  
QY 594 -----TELSNDGTSKDFSVKSGTNNSLVTAEBHLASY--LNEV----- 629  
Db 978 SIGYAGELGNAKATVINGECT--YSL--GNTNLSUTANESGIFGNSNEIKAKENARIV 1031  
QY 630 -NRTADSA-----LQSFVKEEDDDANAITVAKDRTTKNAGAVSILK 670  
Db 1032 GNKNTIGAIBEKPHVPGTTPAAPVNDLKDIIYVTGYDNKISSDKKLAKD-----LSGLF 1084  
QY 671 LKKG-NGLTVATKKDGTIVTGLSQDSGLTIGKSTLNDGL-----TVKDTNE-QIOVG 721  
Db 1085 VYGHGNIAQLPDPDSTEEFTLT-DSSVIGANNLTNKGNYFVLGNVNTATLNSVYLG 1143  
QY 722 ANGKFTNVNPSPGTGIANAR-ITRDKIGFAGSD--GAVDTNKPYPYLDQDKLOVGNVKI 778  
Db 1144 ADS-AYTTGNTSMNSYADMANCLNKGVTTFAGSQPVGVTVGAVGKERRVONVASGLV 1202  
QY 779 TMTGNA-GKATIGLSPTLPSIADQSSRIELGNTTQDKDKSNAASINDILNTGFNL-- 835  
Db 1203 TEASTDAINGSOLFALTRPLRFRAGDNST---LSNP-NGKPGDGTVTVISRSSNOGMKVVG 1257  
QY 836 -KNNNPIDFVSYDVIDVFANGNATTAIVTHDTANKTSKVYVYDVNDVDTTHLTG-TDDN 893  
Db 1258 GENDGNKLTITADKNIGVANGDHTLEVRLAKTLSNLKDATFGTGTDKTINKDGMTITN 1317  
QY 894 KKLGVKTTKLKNTSAN--GNTATNFNVNSDEDAVNAKDI---AENLNTLAKEIHTK 947  
Db 1318 ---GANTVSTLGLNGMGNKNTVNAAGONETDA-VNVRKOLNDLKAEFGGLTGEDGGTVK 1373  
QY 948 ---GTA-----DIALQFTYVKYVDENNADDA--NAITVG--OKNANNQVN-TITLKG 994  
Db 1374 QALGTAIKVTDGDDNVKTKIVTDADGSKKLEIGLENQVTLGGEAKGNPAADGKLTLANQA 1433  
QY 995 GLNIKTDK-----NGTVTGTINTSG-----LKAGKSTLNDGLSLIKNPTSEQIOVG 1042  
Db 1434 G-----TDKVVLDGANGTV--GLTGADGAQAVITVKGRPTLDNAAETPRIAYGNEEVATL 1487  
QY 1043 ADGVKF-----AKVNNN-GVUGA-----GIDGTTIRTRDEI----- 1072  
Db 1488 NDGLKFGANAGDVHNAKINTQVDVKGATVNTVNDNFDKGQINMTRVBGNTITVALAKALS 1547  
QY 1073 -----GFTGNSLDKSKPHLSKDGINAGGKKITNIQSGEIA 1109  
Db 1548 GLTSATFGDPASPKDSTVINKDGLTITQDNTVS---LTDGLDNGNKQIKNVASGLTT 1604  
QY 1110 QNS-----HDV--TGCKIYDLKTELENIKSSTAKTAQNSLHBFVSADCGNFTVSNP 1161  
Db 1605 TNGTATTSLDADVQTVGNVYDGLKTAIINNITNGT-----NPLGGFGLKDRAGNTF----- 1654  
QY 1162 YSSVDTSKTSDVITTFAGENGITTKVNGVVRVJDOTKGLTTPKLTVCNNNGKGVIDSQ 1221  
Db 1655 --KONLCETAQI--TGDSNVNTKV-----VD-----GQNGGKAIEVSL- 1688  
QY 1222 NGQNTITGLSNTLANVNDKSGSVRTTEBQNI IKDEKTRAAISIVDLSAGFNQNGEAV 1281  
Db 1689 --ANQLTLGKPEANVFNATG---EAGKITLKDGDTRVVDVVGSEGAISLTGQPAQ 1741  
QY 1282 DFVSTYTVNFAQGNATKAVTYD-----DTSKTSKVYVDV-NVDDTTI----- 1324  
Db 1742 GAAAPTAKIRVAEGNPDLENTSDDPANPNQNKTRITRYDIAGPNGTVTVEQLATFLNDGLK 1801  
QY 1325 -----EVKDKKLGVK-----TTTTLTGTGTCANKFALSNAQTGDALYKASDIVAHLN 1370

Db 1802 FGANTGVHDAKLNTVRDVVKGAENTWNPADACQ--IMTQISGNTIT-----VALAK 1853  
QY 1371 TLSGDIOTAKGASQAN-----NSAGYVDADGNKVI-----YDSTDNKYYQA----- 1411  
Db 1854 ALAG-LDSATFGNPDGSKDGAVINKDKLTIITEGDKTVKLTKEKGLDNGNOIINVDGLK 1912  
QY 1412 KNDGTVDKTEV-----AKDKLVAQAQTFDGTGLAQ 1441  
Db 1913 KTDGSSVALKDAEGSVLTNGVNVGDLKNAIKDVTSATNGFGFLKDKAGAEPKQDLGTGAQ 1972  
QY 1442 MNVKSVINKEOVNDANKQ-----GIN-EDNAFV----- 1469  
Db 1973 ITGDKNINTKVIDVPNSNDKALEISLANDITLKGNGADGVDSGLGVNGKOGASVVLNGKD 2032  
QY 1470 -----KGLEKAASDNKTKNAAVT-----VGLDNAVA 1495  
Db 2033 GSIGLTPRGDQSDGSKSATISVKDGRAGVDGDKDGTTRIVYETKDATGKPVVEEAVTL 2092  
QY 1496 QTPLTFAAGDTG-TTAKKLGETLTIKGG-----QDPTNKLTONNIG----- 1534  
Db 2093 NDGMKFVNDGKEVTRKLNELTLDIKGGLDAATVADNAKVVSSNLGVKTNAEGTGLEIYMK 2152  
QY 1535 -----VWAGTDG--FTVKLAKD-----LTNLSYNAGSTKIDDK-----GVSVF 1571  
Db 2153 ERPTFSLVWNGDGEDAAVAFKADGDKDGMIAAVTD-NDGNATGLTIKDKDGNPVTFF- 2210  
QY 1572 DSSQAKANTPVLSANGLDLGGKVISNVKGTGKTDTAANVOOLNEVRMLLGLGNAGNDNA 1631  
Db 2211 NNDGR-----ITNVTAGVDDKDAVNVSQLD-----GLAKATTKE 2246  
QY 1632 DGNQVNIA-DIKKDPNSGSSS-----NRTVIKAG-TVLGKGNNDTEK----- 1672  
Db 2247 AGKNMTVPTVNQD--GSTTYTVATEDNVNFTTGTGNTVM---NNDGVKVGDNVALT 2299  
QY 1673 -----LATGSIQVGVDK-DGNANGDLS-----NVWVKTDGSGKKALLATYN 1713  
Db 2300 NEGLKAGDVTVTTAGINAGNKKVTGVADGDISPNSTDAVNGSQLNAVKETAEAGWHHLTA 2359  
QY 1714 AAGOTNYLTNNPABEIDRINEQG-IRFFHVNDGNOEPVVOGRNGIDSSASGKHSVAIGFQ 1772  
Db 2360 GADSSNV---KPRNTVDLNTDGNIVISKTNTADKHNT---FGLADNINVKDSVVVGPK 2413  
QY 1773 AK--ADGEAAVAICRQTA--GNQSIAIGDNAQATGDOSIAI-----GTGNV-----VAG 1818  
Db 2414 GANGKPEGAVVINAEDGANGKDGISI---VGKDGKDAVAISGRKDGVTIGLTPAGADG 2470  
QY 1819 KHSCA-IGDPSTVKADNSYSVGNNNQFTDQTQDVFVGNNITVTESNSVALGSNSAISA 1877  
Db 2471 KNAHAIIGVDSVAGLGDNDGDKGKNSKTRIVYTPNGEEOV-ATMNDGLVFGADKGTGH 2529  
QY 1878 GTHAGTOAK-KSDGTAGTTTTAGATGVTKGPAQTAVGAVSVGAS-----GAERRIQH-- 1929  
Db 2530 KAKLGTTVKVGKDDKNIETEVAGDTIRVR-LKDNIDVKGINVTENLTVKEGAKINMGNNV 2588  
QY 1930 ---VAAGEVSATSDAVNGSOLYKATQSI---ANATNELDRIHONENKANAGISSAMAM 1983  
Db 2589 IDGVADGEVATNSQAVNGSOLHRVQOQVNNQATAINKLGDHINKVKDXDLRAGIAGATAV 2648  
QY 1984 ASMPQAVIPGRSMVTGGIATHINGOGAVAVGLSKLSDNGOWFKINGSADTOGHV--GAAV 2041  
Db 2649 AFLORPNEAGKSVLSGVGSYRSESAIAGVARNSDNNKISIKLGGGMNSRKGVDNFGGSI 2708  
QY 2042 G 2042  
Db 2709 G 2709

RESULT 4

Q9PC04

ID Q9PC04 PRELIMINARY; PRT; 1190 AA.

AC Q9PC04;

DT 01-OCT-2000 (TREMBLRel. 15, Created)

DT 01-OCT-2000 (TREMBLRel. 15, Last sequence update)

DB	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)	348	WNLTASGANSNVA-----LGPSVDLKNSDGNLLITKT-----TDSNDVTFNLATALKV----	396
DB	SURFACE PROTEIN.	439	SSTTAELLSDSLTFPTQNTGQSSTKTVYGVNGVGFNTNAETAAIGTRITRIRDKIGFAR	498
DB	XF1981.	397	-----DSLT-----TG--NTAMTTDGVTVGK-----RVTLDSTG-----	423
DB	Xylella fastidiosa.	499	GDVDEKOAPYLDKOLKLVGSVAITIDNGIDAGNKIINSLAKGSSANDAVTIEOLKAAPK	558
DB	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;	424	-----LVIAEGPSVISSGINAAGQIMNVGTGTADTDAVNFQGOAV-----	465
DB	NCBI_TaxID=2371;	559	TLNAGAGISVTPTEISVDKASGNVTAPTNYIGVTKTELSNDTSKDSFQVANGANGIKFTNVN	618
DB	SEQUENCE FROM N.A.	466	-----SDTASKGWLLASGTNS-----	483
DB	STRAIN=9A5C;	619	AEHLASYLEVNEVRTADSALQSFTVKEEDDDANATVAKDTTKNAGASVILKKGNGIT	678
DB	MEDLINE=20365717; PubMed=10910347;	484	-----NVAPGASVDLKNTDGML	500
DB	RA Simpson A.J.G., Reinach F.C., Artuda P.E., Abreu F.A., Acencio M.,	679	VATKDKG--TVTFGLS-----ODSGLTIGKSTLNDGLTVKDTNEQIQVGANGIKFTNVN	733
DB	RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,	501	LTITKAIGINDVTFNLATALEASLTITGNTAMTDDGVTV-----	540
DB	RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,	734	NPGTGIANTRITRDKIGFASGDVNTKPYLDQDKLOVGNVKTITNTGINAGGKAITGL	793
DB	RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,	541	N-----VT-----LGSTGLVITDGP-----SVTSSGISAGNOKIT--	570
DB	RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa Neto C.M.,	794	SPTLPSIADSSRNIELGNTIQDKDKSNAASINDILNT-----GFLKNNNNNPIDFVSTYDI	850
DB	RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,	571	-----NVAAGTA--DITDAVNFSQLQAVSSTASKGNL-----	600
DB	RA Facincini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,	851	VDFANGNATATVTHDTANKTSKVYVDVNVDDTTIHLTGTDNNKILGVKTKLUNKTSANG	910
DB	RA Fraga J.S., Frasca S.C., Franco M.C., Frohme M., Furlan L.R.,	601	--LASG-----AN--SSNVPGESVD-----LKNSDGN--LLITKTTDSN-----	634
DB	RA Garrier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,	911	NTATNFVNSDDELDALVNAKDIAENLTAKETHYTKGTADTALQFTTVKVVNDENNADD	970
DB	RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kempner E.L., Kitajima J.P.,	635	-----DVTFNLTALUKVDSLTTG-----	657
DB	RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,	971	ANAITVQOKNANNQVNTLTUKGENGLNIKTDKNGTVFFGINTTSGLKAGKSTLNDGLSI	1030
DB	RA Lemos E.G.M., Lemos M.V.F., Lopes S.L., Lopes C.R., Machado J.A.,	658	DGTGVTVG-----SNVTLG--STGL-----VITDG-----	679
DB	RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.B.,	1031	KNPTSGEQIOVGADGVKFAKNNVNVGVGADIGTTRITRDEIGFTGNGSLDKSKPHLSK	1090
DB	RA Marques M.V., Martins E.M.F., Matsukuma A.Y.,	680	-----PSVTS	684
DB	RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,	1091	DGINAGCKITNIQSGEIAQNSHDVATGGKIYDLKTELENKISSATAQNSLHEFSVAD	1150
DB	RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,	685	SGISAGNOKITNVAAGTA--DITDAVNFSQL-----QAVSSTASKGNLL-----	726
DB	RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,	1151	EQGNFTVSNPYSSYDTSKTSVDITTFAGENGITTKVKNKVVVRVGDITKGLTITPLTVGN	1210
DB	RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,	727	ASGANSNVAPGESVDLKNITD-----	749
DB	RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,	1211	NNGKGIIVDSONGONTITGLSNTLANVNDKSVRTEEQNIKDEKTRAASIVDVLSA	1270
DB	RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,	750	-----IVISKESSN-----	762
DB	RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,	1271	GFNLQNGEAVDFYSTYDITVNFADGNATTAKVTYDDTSKTSKVYVDVNVDDTTEVDRKK	1330
DB	RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,	763	-FNLSS-----	773
DB	RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,	1331	LGVTITLTSTGTGANKFALSNOATGDALVKASDIVAHLNTLSGDIQTAKGASQANNSAG	1390
DB	RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,	774	LTVGDTVMTTNG-----	785
DB	RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,	1391	VYDAGNKVIYDSTDNKYYQAKNDCTVDKTEKAVDKLVAQAQTPDGTLAQNNVKSINK	1450
DB	RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.,	786	-----	785
DB	RA "The genome sequence of the plant pathogen Xylella fastidiosa."	1451	BOVDANKKQGINEDNAFVKLEKASDNKTKNAATVYDGLNVAQTPITPTFAGDTGTAK	1510
DB	Nature 406:151-159(2000).	786	-----SOVTLG	795
DB	EMBL: AE004017; AAF84783.1; -.			
DB	SEQUENCE 1190 AA; 118446 MW; 756741B0C8D787CC CRC64;			
DB	Query Match 7.6%; Score 778; DB 2; Length 1190;			
DB	Best Local Similarity 20.5%; Pred. No. 8.1e-25;			
DB	Matches 429; Conservative 185; Mismatches 432; Indels 1044; Gaps 72;			
DB	64 SAYAOKDKTHIAICEQNPQRS--TAKAD-----GDRA 96			
DB	39 SAASSKCT-----QPRSNNAATKRSALKHRRQLHVLTLVLAASATGYTGKVA 89			
DB	97 IAGENNAQGGQAIAGSSNKT--VNGSSLDKTKGTATGQE-----SIATGGDVKA 146			
DB	90 AQVYVNSDSTENCVEILLGDSSTQSFHISASNDCKPDPOTFEYSLFYDYNLVLG-- 145			
DB	147 SGDASIAIGSDLLHLDQGNPKPK--GTLNDLNLINGHVLKEIRSSKDNDDVYKRRITAS 205			
DB	146 ----SLYVNEKGLGLVDSIGATYSKRLGSIAT--MNGSAGIDSI-----AIGS 187			
DB	206 GHASTAVGAMS----YAQGHFNAFCTRATAKSAYSLAVGLAATAGQSTIAIGSDATSSS 262			
DB	188 GQSKTDGNTSGATVAQGLRSLAIGTITARSQSDAISIGTGASTTGNFAIAGNALTISI 247			
DB	263 LGATLGAAGTQAQGLGSGVVTGSDNNSRPAYTPNTQALDPKFOATNNTKAGPLS 322			
DB	248 ANGIALGASSSVTTGGVALGOGSLLAATASGIT--GYDPVTKSTSTLSTSMWRSTLGAVS 305			
DB	323 IG----SNSIKRKIINVAGVKNKTDVAVNVAQLEAVVVKAKERRIFPQGDNDSTDVKIGLD 378			
DB	306 IGNITSSTQTRQLTGLAAGRSDDTDAVNVVAQLKLAE-----SVGGG 347			

QY	1511	KLGETLTIKGGQTDNNKLTDDNIGVVAGTDGFTVKLAKDLTNLNSVNAGTFKIDDKGVSF	1570
Db	796	SMGLVI-----TDG-----	804
QY	1571	VDSSGQAKANTPVLSSANGIDLGGKVISNVGKGTKTDTDAAANVOOLNEVRNLLGLGNACNDN	1630
Db	805	-----PSVTSNGINAGSQKITVWAAGTADTDADVNLSQLN-----	838
QY	1631	ADGNQVNIADIKKDPNSGSSNRTVIKAGTVLGGKGNNDTEKLAT--GGIQGVGDKDQNA	1688
Db	839	-----TAMASGAKSVHYIYTYDGGTQ-----GGNY	864
QY	1689	NGDLSNVVWVKTDKDGSKKALLATYNAAQTNYLTNNPAEIDRINEQCIFFHFVHNDNQE	1748
Db	865	NGD-----	867
QY	1749	PVVOGRNGIDSSASGKHVSVAIGFQAKADEAAVATGRTQTA-GNOSIAIGDNAOATQDQS	1807
Db	868	-----GATGTRSIAVGVGTLASAEGATAVGSAAASGKSTAIGRNAVASADGS	916
QY	1808	IAITGC-----NVVAGKHSG-----ATGDPSTVKADNSYSVGNNNQFTDAT--	1848
Db	917	VALGDGAKDCARGAESYTKYSGLQNNVTGTVSGDASKGTRTVTSNVAOAKEATDAVNL	976
QY	1849	-QTDVFC-----VGNNI-----TVTESNSVALGSNSA--ISACTHA-----GTQAKK	1887
Db	977	RQLDRVAQDANRYVDNKIESLSEGGTFVKVNSL---NNSATPIAAGVDATAIGVATPAG	1033
QY	1888	SDGTA-GTUTTTAGATCTVKGFAGQTAVG-----AVSVGASGACRRIONVAAGEVSA	1937
Db	1034	ADSIAMCNKASADNAV-----AIGNHVSADRANTVTVSGSAGSERQVTVNVAAG---T	1083
QY	1938	TSTDVANGSOLYKATQSIANATNELDHRHQNENKANAGSISSAMAMASMPQAYIPGRSMV	1997
Db	1084	ADTDVAVNSQL---NQGLITAKQYTDGVVGSILRRDTDGVAATAATANLPOAYIPGRGMT	1140
QY	1998	TGGIATHNGCGAVAVGLSKLSNDGOWPFKINGSADTOGHVGAAVGAEHFH	2047
Db	1141	SVGVSSYRQGSIAIVGVSSVSGRWPFKFGSNTSRQSVQIGVAGVQYW	1190

## 5 RESULT

Q9F3X0 ID 09F3X6 PRELIMINARY: PRT: 1299 AA.

AC	Q9FX36;	
AD	DT	01-MAR-2001 (TrEMBLrel. 16, Created)
AE	DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
AF	DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
AG	DT	MAPA PROTEIN.
AN	GN	MAPA.
AO	OS	Pasteurella multocida.
AP	OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
AQ	OC	Pasteurella.
AR	OX	NCBI_TaxId=747;
AS	OX	[1]
AT	RN	SEQUENCE FROM N.A.
AV	RP	STRAIN=PW70;
AW	RC	Henderson I.R., Nataro J.P., Cappello R., Stein C.;
AX	RT	"Evolutionary origins of the autotransporter proteins.;"
AY	RL	Submitted (APR-2000) to the EMBL/Genbank/DDBJ databases.
AZ	RL	EMBL; AJ277635; CAC14202.1; -.
BA	SQ	SEQUENCE 1299 AA; 130963 MW; 8BCCE0EB6CDB428 CRC64;

Query Match 7.2%; Score 739; DB 2; Length 1299;  
Best Local Similarity 20.8%; Pred. No. 3.9e-23;  
Matches 443; Conservative 231; Mismatches 545; Indels 908; Gaps 84;

QY	1	MNHIYKVF	KNKATG	TFMAVA	YAKSH	STGG	SCATGQ	---	VGS	CVTL	SFAR	TAAL	VLVI	57
Db	1	MNKIYPTL	WNAATQ	SWVVV	SELAKA	---	GCK	SACK	SALV	NSVCS	FTLIA	ASVWL	--	54

Qy	.58	GATLSSGAYAQKDXDTKHIAIGBONQPRRGSTAKACGDRAIRAIGENANAGGOQAIGA	SSN 117
Db	55	-----GS-----GOVNAAEVTGNTGVSGDDKYCF--YNASSQSVICGDATT	93
Qy	118	KTVNGSSLDKIGTDANGQESIAIGDGVKASGDASTAIGSDDLHLLDHGNPKHPKGTGLN	177
Db	94	KT-----TDKT-----	99
Qy	178	DLINGHAVLKEIRSSKDNDVKYRRTTASGHASTAVGAMSYAOGHFNSAFGRATAKA	SAYS 237
Db	100	-----DN-----	KPAKS 106
Qy	238	LAVGLAATABEQGTIIAGSDATSSSUGAIALGAGTRAQLQGSIALGQGSVVYTQSDNN	-- 295
Db	107	VWIFGATNDGETNVAIGAKSKSAASTAIGDNAKALDNOAIAIQONATANSDWDISIG	166
Qy	296	RPAYTPNQALDPKFQONTNRKAGPLSIGSNSIKRKILNVGAVNKTDVNVQOLEAVVK	355
Db	167	RQAGEQTEV-----SABERNTAIDGGALKR-----GKGVN-	199
Qy	356	WAKERITFOGDDNSTDVKIGLDNLTLTKGAETNALTDNNIVGVYKEADNSGLKVKLAKT	415
Db	198	-----NNIALTSGAD-----RLAGT	213
Qy	416	LNNLTEVNTFLNATITVTVKG-SSSTTTAELLSDLSTFTQPWTGQSOSTKITVYGVMGKVF	474
Db	214	HNVLM--GYVNDAEAVRSALTAGSTTKEI-----NAKETTDNKYYI----	253
Qy	475	TNNAETTAIGT---TRITROKIGFARGDVEKQAPYLKKOLKV-GSVAITIDNGID	529
Db	254	--EASNVTALUGTRALATQAAVAIG-----OQAKAFGNSVAVGNSTK 294	
Qy	530	AGHKIKSNLAAGSSANDAVTI EOLKAAKPPTLNAGAGISVTPTEISVDASKGNVTAPTNYI	589
Db	295	ASQG--TAITAGSNAH-----ATGSSSIAIGTGVNGDNTARTLASDNFTA--M 339	
Qy	590	GVKTELNSDGTSDKFSVGSGTGNNSLVTAEHLSVLNEVPNTADOSLOSFTVKEEDDDD	649
Db	340	GLSQAQTKSDAIAVGNNAAGINTYSI-----GYNAGVTKTAD-----EAQSDP 384	
Qy	650	ANAITVAKDPT--KNAGAVSILKLGKNGLTVAITKKDGTVTFFGLSQDSGLTGICKSTLNN	706
Db	385	NNKLEPATDAVFIGNKAGYKS-----NONRMQVSJLKDSGE--GVVGTEENTYIGNSACKN	437
Qy	707	DGLTVTDTEBQIOVGANGIKFTNVNGSNPGTGTANTARITRDKIGFAGSDG-----	757
Db	438	--TKGNTVNAISRAG---QNVECHD-----NFAALIENGONIKESDNIAITCKHNAGR	484
Qy	758	AVDTNPKYLDQDKLOYG--NVKITNTGFINAGKAITGISPLTLPISIADOSSRNIEL--GNT	813
Db	485	SADPNTKLINNT ISLGKESVSLKNFGIAQGNKAKT---DGLASIA--IGRNAEAVGGET	539
Qy	814	IDOKDKSNAASINDILNTGFNLKNNNPIDFVSTDVIDVFANGNATTATVTHDTANKTSK	873
Db	540	-----ANIA-----IGDSASADASGAIVLGTQAQAKSL 567	
Qy	874	VYDYVNVDDTTIHLTGTDNNKLGKVTTKLNTKTSANG-NTATNFNVNSSDEDAL-VNAKD	931
Db	568	TV-----DGKKYGAYSIVIGTEKATAQAAPACKNPNKPDIAIATGKA 611	
Qy	932	IAENLNTLAKEIHHTKTADTALOTFTVKVVDENNADDANAITYGO--KNANNOVNTLT	989
Db	612	EAHVASTIALGFCAKSDTKAQAV-----SIGYNSNAKGYQAIAFGSEAKTTENAGSSIA	665
Qy	990	LKGENGLNIKTDKNGVTTFGINTTSLKAGKSTLNDGGLSIKNPTGFSBOIQVG-----	1042
Db	666	F---CTKAQTRASASIAICMGAEETF-----DGGQAL---DGSDAVALGREAKAKR	710
Qy	1043	----ADGVKFAKVNNGV-VGAG-----IDGTRITRDEIGFTGTNGSIDLSKPHLSKDG	1092
Db	711	ONALATFYKAVADHKQAVALGAGETAAGAEGTEATVNEFKTYSYFRAG-----IKPAT---	763
Qy	1093	INAGGK----KITNIOSGEIAQNSHDVATGCGKIYDILKTELENKISSTAKTAQNSLHFPSV	1148

Db 764 VSVGKDAERTITNVAAGRIDKSTDAINGSOLY-----LALNAL----- 803  
 Qy 1149 ADEQGNFVSNPSSYDFSKTSVDITFAGENGITTKVKNKGVVRGIDQTKGLTPKLV 1208  
 Db 804 ----- 803  
 Qy 1209 GNNKRGIVDSONGQNTITGLSNTLANVTNDKGSVRTTEQNIIDKEDKTRAASIVDVL 1268  
 Db 804 GN-----VGNLT--VTNVLGG-----DAIVKEGDE----- 827  
 Qy 1269 SAGFNQGNCEAVFVSTVDYTNFADGNATTAKVIYDDTSKTSKVVDVNVVDVDTIEVKD 1328  
 Db 828 -AG----- 829  
 Qy 1329 KKLGVKTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSGDIQTAKAGSOANNS 1388  
 Db 830 -TLTMSNIGG-----TGKG----- 842  
 Qy 1389 AGYVDAGNKVYDSDTDNKYIQAQNDGTVDTKEVAKDKLVAQAQTPDCTLAQMNKSVI 1448  
 Db 843 -----TIHDA----- 852  
 Qy 1449 NKEQVNDANKKOGINEDNAPVKGLEKASDNKTKNAAVTVGDLNVAQAQPLTFAGDTGTT 1508  
 Db 853 NTAASKTTVEG---DNITV--TEEAAD----- 877  
 Qy 1509 AKKLGETLTIKGGQDTNKLTDNNGVAVAGTGFYKLAKDLTNLSNVNAGTKIDDKGV 1568  
 Db 878 -----GSRVTYVATKDDV-KFDSVAVAGTKIDANGL 907  
 Qy 1569 SFVDSSGQAKANTPVLSANGILGKVLISNVKG--TKDT--DAANVQOL-----NEVRNLL 1621  
 Db 908 TFVDDQGTIDNTPSISKTGIDAGNQKVTWQNGNIADKSDAVNGQOLFAOGEVKNII 967  
 Qy 1622 GLGNAGNDNADGNQVNIADIKKDPNSGSSNRVTVIKAGTVLGGKGNNDTEKLATGGIQVG 1681  
 Db 968 G-----GDTTYNPETGEYAN-----TNIGGTGASTI----- 993  
 Qy 1682 VDKGNANGDLSNVVTKQDGSKKALLATYNAAAGOTNLNTPAEADRINEQGRFFH 1741  
 Db 994 -----DEAKVNTAKAA----- 1007  
 Qy 1742 VNDGHOEPVVGNGRIDSSAKSHSVAIGFOAKADGEAAVAGROTQAGNOSIAIGNAQ 1801  
 Db 1008 -----KTEVQGENIVVTSAPCAN-----GNTVYTVATAKE 1038  
 Qy 1802 ATGDOSIAIGTGNVYAGKHSAGIDPS--TVKADNSYSGVN--NNQFTDQTOTDVFVG-- 1855  
 Db 1039 VTDFDKTT---VGSVYTDKNNDITGLSKNTLGGDNFAKNGRAASEEQLNATQTNLATILG 1095  
 Qy 1856 GN-----NITVTESNSVALGS--NSAISAGTHAGTQAKK-----SDGTAGTITTAGATGT 1903  
 Db 1096 GNAQNTNGNVAMTDIGGTCNNINDAIKASRNEVKQGNMNVVPTPTTGANGQTIYEATAD 1155  
 Qy 1904 VKFAGQTAAGVAVSGCAERIQNVAAGEYSATSTDAVNGSOLYKATOSIANA---TN 1960  
 Db 1156 KVAF-DEVKVGITIDAT--TKISGIAKGDISENSTDVNGSOLYELFQOKIARSGDNYN 1212  
 Qy 1961 ELDRHTONENKANAGISAMAMASMPQAYIPGRSMVTGGIATHNGQAGAVGLSKLSDN 2020  
 Db 1213 ILNRRINKVDKDLRAGIACANAAAGLPQAYIPGKSMVAAGTYKQNAIALGMSRISDN 1272  
 Qy 2021 GOWFKINGSADTQGHVGAAGVAGFH 2047  
 Db 1273 GKVIKILKLTGNTNSRGDFGAGYQW 1299

RESULT 6  
 Q48152 Q48152 PRELIMINARY; PRT: 1098 AA.  
 AC Q48152;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE ADHESIN (H1A).  
 GN H1A.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NONTYPEABLE STRAIN 11;  
 RX MEDLINE=96332658; PubMed=8730864;  
 RA Barenkamp S.J., St Gene J.W. III;  
 RA "Identification of a second family of high-molecular-weight adhesion  
 RT proteins expressed by non-typable Haemophilus influenzae.";  
 RL Mol. Microbiol. 19:1215-1223(1996).  
 DR EMBL: U38617; AAC43721.1;  
 SQ SEQUENCE 1098 AA; 114100 MW; D977335A89F7333D CRC64;  
 Query Match 6.6%; Score 678.5; DB 2; Length 1098;  
 Best Local Similarity 23.6%; Pred. No. 1e-20;  
 Matches 304; Conservative 157; Mismatches 428; Indels 399; Gaps 50;  
 Qy 944 HTTKCTADTA---LQTFVTKVVDENNADANAI-TVGQKNANNQVNTL-----TL 990  
 Db 26 HTKASATVAVAVLATLLSATVEANNTPVTKLKAYGDANFNFTNSIADAEEKQVQEA 85  
 Qy 991 KGENLINIKTDKNGVTVEGINTTSGLKAGKSTLNDGGLSIKNPTGSGIQVAGDGVKFAK 1050  
 Db 86 KLLNLLN---EKNASDKLLVEDNTAAATVGNLRKLGWLVSSKNGTRNEKSO----QVKAH- 137  
 Qy 1051 VNNNGVVGAGIDGTTTRTDEIGFTGTNGSLDKSKPHLSKDGINAGGKKITNIQSGEIAQ 1110  
 Db 138 -----DEVLFEG-----KGVQVTS----- 152  
 Qy 1111 NSHDAVTGCKIYDLKTELENKISSTAKTAQNSLH--EFSVADEQGNNTVSNPYSSYDTS 1168  
 Db 153 -----TSENGKHTITTEALAKDLG-----VKTA 174  
 Qy 1169 KTSVDVITFAG--ENGITTKVKNKGVVRGIDQTKGLTTPKLV--GNNNGKGVIDS--QNG 1223  
 Db 175 TVSDTLTIGGGAAGAT-----TTPKVNVTSTTDLKPAKDAAGANG 216  
 Qy 1224 QNTI--TGLSNTLANVTNDKGSVRTTEQNIIDKEDK--TRAASIVDLSAGFNQ----- 1275  
 Db 217 DITVHLNGIGSLTIDL--VGSFATHIDGG---DQSTHYTRAASIKDVLNAGNINIKVKA 271  
 Qy 1276 ----GNGEAVDVFSTVDYTNFADGNATTAKVTYDDTSKTSKVVDVNVVDVDTIEVKDKKL 1331  
 Db 272 GSTTGQSENVDVFHTYDTVEFLSADTETTTVTVDSEKNGKRTVEKIGAKTSVIKEDGKL 331  
 Qy 1332 GV-KTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSGDIQTAKAGSQAANSAG 1390  
 Db 332 FTGKANKETNKVDGN--ATEDADEGKGLVTAKDVIDAVNKTGWRIKTTDANGNGDFAT 389  
 Qy 1391 YVD-----ADGN---KVIYDSDTNKYYQAQNDGTVDTKEVAKDKLVAQAQTPDGTGLAQ 1441  
 Db 390 VASGNTVTFASNGTGTATVTNGTDG--ITVYDAKYDAGDGLKIDGDKIAA-----DTALT 442  
 Qy 1442 MNVKSVINKEQVNDANKKOGINEDNAPVKGLEKASDNK---TKNAAVTVGDLNVAQAQTP 1498  
 Db 443 VN-----DGKNANNPKGVADVASTDEKKLVTAKGLVTA--LNSLSWT 484  
 Qy 1499 LTFAGDTGT-----TAKKLGTELTIKGGQDTNKLTDNNGVAVAGTGTGFTVKLAKDLT 1551  
 Db 485 TAAEADGDTLDGNASEQEVKAGDKVTEFKAG-----NLKVKQBGANFTLSQDALT 535  
 Qy 1552 NLNSV-----NAGTKIDDKGVSEFVSSGOA--KANTPVLSANGLDLGKVISNVKGT 1603  
 Db 536 GLTSITLTGTGNGAKTEINKDGLIITPANGAGANNANTISVTKDGISAGSGSVKNVYVSG 595  
 Qy 1604 KDTDAANVQOL-----NEVRNLLGLGNAGNDN-----ADGNQVNIAD----- 1640



```
Db 596 KFGDANFDPLTSSADNLTKQNDAYKGLNLDKGTDKQTPPVADNATAATVGDRLGLW 655
QY 1641 -IRKDPNSGSS-----NRTVIKAGTVLGGKGNNDTEKLTGCGIQGVGDKDGNANGDL 1692
Db 656 VISADKTTGSGTEYHDOVRANEV--FKSGNGINVSQKTVNGRREITPEL---AKGEV 709
QY 1693 --SNVWVKTKQSGKALLAT---YNAAGTNYLTNNPABADRINEQGRFFHVNDGNQ 1747
Db 710 VKSNEFTVKETNGKTSLVKVGDKYYSKEDIDLTGP-----KLKDN- 753
QY 1748 EPVVOGRNGIDSSASGKHSVAIGFOAKADGAAVAIGRTQAGNQSTAIGDNAQATGDQS 1807
Db 754 -----TVAAYQDK-----GKVVSVVDTEATITNK 780
QY 1808 IAITGTVNAGKHSAGIDPSTVKADNSYSVGNNO-----FTDAQTQDFVGVGNITV 1861
Db 781 --GSGVYTGQVADAI-----AKSGFELGLADEADAKRAFDDTKALSAGTEIVNA 830
QY 1862 TESNSVALGNSAISAGTHAGTQAK----- 1886
Db 831 HDKVRFRANGLNTRKVSAAATVESTDANGDKVTTTFVKTQVDELPLTQIYNTDANGKKITKVVK 890
QY 1887 -----KSDGTAGTTT-----AGATGTVKGFAGQTAVG 1914
Db 891 DGQTKVELNADGTADMTKEVTLGNDVSDGKKVVKNDGKWKYHAKADGTADTKGVEVSD 950
QY 1915 AVSVG-----ASGAERRIQNVAAGEVSATSTDAVNSQLY---KATQSIANAT 1959
Db 951 KVSTDEKHVYSLDPNDQSKGVVDNVANGDISATSDAINGSQLYAVAKGVYTNLAGOV 1010
QY 1960 NELDHRIHONENKANAGISSAMAMASQPAYIPGRSMVTGGIATHNGQGAVALGLSKLD 2019
Db 1011 NNLEGVKNGYKGRADAGTALAASQLPQATMPCKSMVAITAGSSYQOGNGLAICVSRISD 1070
QY 2020 NGQWFKINGSADTQGHVGAAGAGRHF 2047
Db 1071 NGKVIIRLSTGTSQGTGVAAGVGYOW 1098

RESULT 7
Q9F2D8 PRELIMINARY; PRT; 1107 AA.
ID Q9F2D8
AC Q9F2D8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE SAPB PROTEIN.
GN SAPB.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CF18;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ277623; CAC14217.1; -.
SQ SEQUENCE 1107 AA; 113223 MW; F5C7CA651FED51AB CRC64;

Query Match 6.4%; Score 661; DB 2; Length 1107;
Best Local Similarity 23.3%; Pred. No. 5.6e-20;
Matches 298; Conservative 195; Mismatches 418; Indels 368; Gaps 58;

QY 924 DALVNA-----KQIAEN--INTLAKEIHTTKGTA---DTALQTFYVKKVDENN 966
Db 42 DALANAGNDTGSINQNTDTLATNTTSNNLSNVTTLTDDALLWDAASGTFASR---NG 98
QY 967 NADDANAITYGQRNANNQVNTLTILKEN--GLNKTDKNGTVFGINTTSLKAGKSTLN 1024
Db 969 VNSVLNLGDSGGTTRIGNVSAAVNDTDVAVNQLKRSVEANTYTDQKMGEMNSIKG 1028
```

```
Db 99 SASKITNLAAGTLAAD---STDVANGSQLFDTNEKVDQN-----TADITNTNNSIN 146
QY 1025 DGLSINKPTCSQIQVAGDCVFKFKNVNGVVGAGIDGTRITRDEIGTGTNGSLDKS 1084
Db 147 QNTIDTATNTS-----INNLS-----NSVTTLTDDALLWDAASGTFAS 186
QY 1085 KPHLSKDGGINAGKKITNIOGGEIAQNSHDVATGGKIYDLKTELENKISSTAKTAQNSLH 1144
Db 187 R-----NGSASKITNLAAGTLAADSTDAVNGSQLFDTNEKVDQN--TADITNTNNSIN 237
QY 1145 EFSVADSGGNFTVSNPYSYDTSKTSQVI-----TE-AGENGITTKV-NKCVVRVGI 1195
Db 238 QNTT--DIANTNTSINNLSNVTTLTDDALLWDAASGTFASRNGSASKITNLAAGTLAA 295
QY 1196 DOTKGLTTPKLTGVNNNGKIVIDSQN-QONT--ITGLSNTLANVTNDKGSVSRTEQONI 1252
Db 296 DSTDAV-----NGSQLYETNQVDQNTSAIDINTSITNLSDD----- 333
QY 1253 IKDEDKTRAASIVDVLISAGNLOQNGEAVFVSTYDVFNFADGNATTAKTAYDDTSTKSK 1312
Db 334 -----NLSWN-----ETNFSFASHGSSSTTKITNVAAGELSE 366
QY 1313 VVYDVNVDDTTIEVKDKKLGKVTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLT 1372
Db 367 ESTDA-VNGSQLFETNEKVDQNTTIDIAANTN-----ITQNSTALENNTS 411
QY 1373 SGDIQTAGKSAQNNNSAGYVDADGNKVIYDSTDNKYIQAQNDGTVDRKTEYAKDKLVAQA 1432
Db 412 VSDINT-----SITGLTD--NALLWDB-DTGAFSANHGGSTKITNVA-----AGA 454
QY 1433 QTPDGTIAQMNKSVINKEQVNDANKKQGNEDNAFVKGLEKAASDNKTKNAAVTVGDNL 1492
Db 455 LSEDST-----DAVNGSQLYETNQK-----VDQNTS-----AIADIN 486
QY 1493 AVAQTPLTEAGDTTAKKLGETLTIKGGOTDTNKLTDNNITGVVAGTGDFTVKLAKDLTN 1552
Db 487 ----TSTINLTGTDALSWDDEGAFSASHGTSGTNKIIN----VAAG-----ELASDT- 531
QY 1553 LNSVNAGGTKIDDKGV---SFVDSGQAKANT--PVLISANGLDLGGKVISNVGKGTDTD 1607
Db 532 -DAIN--GSQLYETNMLISQYNESISQLAGDTSETYITENG--TGVKYIRTDNGLQEQD 586
QY 1608 A-----ANVQOLNEVRNLLGLGNAGNDNAGNQVNIADIKKDPNSGSSSRTVIKA--- 1658
Db 587 AYATNGATAVGYDAVASGAGCALGQNSSSSIEGSA-----LGSGSTSNRAITTGIRE 641
QY 1659 -----QTVLGGKGNNDTEKLTGCIQVGVVDKNGNDLSNVVWVKQKQSKKALLATY 1712
Db 642 TSATSDGWI---GYNTDRELLGALSIGT--DGESTYQITNV-----ADGSE----- 684
QY 1713 NAAQTNYLTNNPAEADIRNEQGIREFVHNDGNQEPVVGQNRNGIDSSASGKHSVAIGFQ 1772
Db 685 ---AQDAVTVRQLQNAICAVTTTPTKYHANSTEE-----DSLAVGTDLSLWCAK 731
QY 1773 AKADGEAAVAIGRTQA---GNQSIAGDNRQAQATGDSIAIGTCNVVNGKHSIGAIDPST 1829
Db 732 TIVNADAGIGIGLNTLVMDAINGIAIGSNARANHANSIAMNGSOTT---RGAQTDYTA 788
QY 1830 VKAD-----NSYSVGNNN---QFTD---ATQTDVFCVG-----NNITVTESN 1865
Db 789 YNMDTPONSVEFVSQVSGEDGQRTNVNAGSADTDVNVGQLKVTDAQVSRNTQSITNLN 848
QY 1866 -----SVALGNSAISAGT---HAGTQAKSKDQTA-GTTTTAGATGTVKGFAGQT 1911
Db 849 TVQSNLDRVTNIEINGIDIVTTGSKYFKTNTDGAADANAOGADVAIGSGSIAAENS 908
QY 1912 AVG-----AVSAGSAGERRIQNVAA----- 1932
Db 909 ALGNTSVADEANTVSGVSTQQRRTITNVAAGVNNNTDANVNAQLKASEAGSVRYETNADGS 968
QY 1933 -----GEVSA--TSTDVANGSOLYKATOSIANATN---ELDHRIHQ 1968
Db 969 VNSVLNLGDSGGTTRIGNVSAAVNDTDVAVNQLKRSVEANTYTDQKMGEMNSIKG 1028
```

Thu Sep 13 14:19:01 2001

QY	1969	NENKANAGISSAMAMPOAYIPCRSHVTTGGIATHHQGAVAVGLSKLSDNGQWFKIN	2028
DB	1029	TENKMSGGIASAMAMAGLPQAYAFGANNTSIAGTFFNGESAIVGSMVSESGGWYKLG	1088
QY	2029	GSADTQGHVGAAGVAGGFHF	2047
DB	1089	GTSNSQGDYSAAIGAGFOW	1107
RESULT	8		
Q9F285			
ID	Q9F285	PRELIMINARY:	PRT: 3705 AA.
AC	Q9F285		
DT	01-MAR-2001	(TrEMBLrel. 16, Created)	
DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)	
DT	01-MAR-2001	(TrEMBLrel. 16, Last annotation update)	
DE	YAPH PROTEIN.		
GN	YAPH.		
OS	Yersinia pestis.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Yersinia.		
OX	NCBI_TaxID=632;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRATN=CO-92 BIOVAR ORIENTALIS;		
RA	Henderson I.R., Nataro J.P., Cappello R., Stein C.;		
RT	"Evolutionary origins of the autotransporter proteins.;"		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AJ277631; CAC14227.1; -		
SQ	SEQUENCE 3705 AA; 370756 MW; 714PDF16455968C9 CRC64;		
Query Match	6.1%;	Score 627;	DB 2; Length 3705;
Best Local Similarity	21.1%;	Pred. No. 7.4e-18;	
Matches 487;	Conservative 329;	Mismatches 876;	Indels 614; Gaps 107;
QY	1	MNHYKVIENKATCTFMVAEYAKSH---STGGGSCATG-----	36
DB	1	MNTFKVIWASLNWVWVVELAKGRKTKSSRNLLISEGLVPKFEQSMVSKFRKNLLAL	60
QY	37	QVGSVCTLSFARIAALAVL-----IGATLGSAYAKQKDKTHI-----	85
DB	61	SLGSLVFLSTGPVPAADITVSTQELSAALSNGTY-----DKILLGADITLGLSVNMT	115
QY	86	SGTAKADGDRATIAIGENANAQGGQAIAGSSNKTV-NGSSLDKIGIDATQESIAIGDV	144
DB	116	SNQVVIDQGGFGLTVNNTNYGLVSSGSGTLTLQNMKID----SANYYSMVVLNG--	169
QY	145	KASGDASTAI-----GSDDLHLDDHG-----NPKHPKGTLLINDLINGHAYLKEIRS	191
DB	170	-----ANTAVNVIYNNIDFLGSSQLIYMGAYGAATNSIMTFGDILND-----VYVNDRA	218
QY	192	SKDNDVKYRRYTSCHASTAVGAMSYAOGHFSNAGFTGRATAKSAYSLAVGLAATAEGOST	251
DB	219	OBIGEVNKLATGRPHVTHTGSSVT-----SFVSTGGANNSTMTDFASGADVKIDRTGST-	273
QY	252	IAIGSDATSSSLGATLALGAGTAAQLOGSIALGQGSVVTQSDNNRPAITPNTQALDKFQ	311
DB	274	-----GDLTSGVNAFAY-----TFADGASPELIANQN-----VFS	304
QY	312	ATNNYKAGPLSTGS-NSIKRKTIINVGAGVKNYTDVAVNAQLEAVVWKAKERRITFQG-DDN	369
DB	305	GTTTNRG--LEIGSYNSID-----GFGSGV-----KIVLQSRSDG	337
QY	370	STDVKTIGLNTLTIKGGAETNALTNNI-----GVKKEADNSGLVKLAKTLNLTNFEVN	423
DB	338	SIISGNCIDNATNAAGINNAGSDANVIYNLGTGSILKATNTG--ILATKNANASDIY	395
QY	424	-----TTTLNATTVKV-----CSSSTTAELLSLSTPQNTGTSQTSKTVY	467
DB	396	IRSAGDITAAATGISAHNGTGTVKIKNDGTITSTTATGAISSASIKSEISVDNTDGTITAT	455

QY	468	GVNGVKFTNNAETTAAGT--TRITRDKIGFARDGDVDEKQAPYLDKKQKLVGVSVAITID	525
DB	456	AGTVNVLASAILNLFGGTINTSATANGITEA-----GTEGGHTLTDLTINLLGTGIALS	510
QY	526	NGIDAGNKKISNLA-----KGSSAND--AVTIEQLKAAKPTLN---AGAGISVTPTEISV-	575
DB	511	N-VAGVNLTLSNVTLNLTNGTALNSLTGLTLDVSLNGRNTINIEGAGIGIAATNTLWTF	569
QY	576	DAKSGNVTAPTYNIGVKTTELNSDGTSDKFSVKSGSTGNSNLSVTAIEHLASLYNEVNRADS	635
DB	570	DAEALDINVNGAGIGIQAT-----GGGVNLS-----ASNLIINVAINTLGT	609
QY	636	ALQSFVTVKEEDDDANAIIVAKDTTKNAGAVSILKLGKNGLTV---ATKK-----	683
DB	610	ALQI-----TDGIDNTTITIGNEIQLNAENATAINFILGSSSKTLLNNGTIRKISVPAGVA	663
QY	684	-----DGTVTFGLSQDSLITIGKSTFLNNDGTLVKTDEIQIOVGANGIKFTNVNGS	733
DB	664	DHINNNGTLDGTLTTCAGNDT-LVIDSSOSNDVINLGDGNSVTIQNGATVSSIITCN	722
QY	734	NPGTGIANTARITRDKIGFAGSDGAVTNKPYLDQDKLVQGVNKKIINTGINAGGKAITGL	793
DB	723	GNDTFTINGMSVSGTYLGLSDAGTGLNTXNXASTDELAAATSLQGTINLNVDSHITLV	782
QY	794	SPTLPSIADQSSRNIELGNTIQDKKSNAASINDILNTGENLKNNNNPIDFVSTYD-IVD	852
DB	783	S-----DDNTGSGMWNIDSSSELFGSTFGDILH	811
QY	853	FANGNATTATVHTDANKTSKVYDVNVDDTTHLTGTDGDKKLGKVTTKLNTKSANGNT	912
DB	812	ATLCAGTGSALVNNSAN-----VSEQASM-FACTWQVNOGQALITASNSQLGSAKI	862
QY	913	ATNPNVNSDEEDALVN-----AKDIAENLTLAKETHTTKGTADTALOTFTVKKVDEN	965
DB	863	GIDGTLN-LDNIALFNHVLVTGTLNVAKNLATTPDFGSGTVGAFSGIVNLT--KITFA	919
QY	966	NNADDANAI-TVQOKNANNQVNTL--TLKGENGLINIKTDKNGTVTFGINTSGLKAKGST	1022
DB	920	LSADNAAALASATLKLSDSVTVTGTDTLHGLDL---SGGLTFDGAVPQOSTGVT	976
QY	1023	LNDGSLSTKNTPTGSEQIOVGADGVKFAKVNNGVVGAGIDGTTTRTDEIGFTGINGS-L	1081
DB	977	VTD--LALNSGT-----VN---ITGSG-----SWDNTDPLATNVISIL	1008
QY	1082	DKSKPHLSKOGINAGGKKTNIQSGEIAQNSHDAVTGKIYDLDKTE-LENKISSTAKTAQ	1140
DB	1009	EODRAGSTLELINA-----TNV-TGDI--DALDLVNGTALTSGTQGVQSAIQGGSTVA	1060
QY	1141	NSLHEFSVADEQGNFTVSNPYSSYDTSKT-----SDVITFACENGITTK--VNKGVVR	1192
DB	1061	NAIHNYGLASSNSNG--DSGLVNVYTLTSALELLADGADALLATESGLTANRVLNAELFG	1118
QY	1193	VGIDQTKGLTTPKLTVGNNNGKIVIDSONGQNTITGLSNTLANVTNDKGSVRTTEQONI	1252
DB	1119	VG-----GLVVDAQNG-----ALTLANGSNRYEGTIVTAGEL	1151
QY	1253	IKDEDKT-RAASIVDVLVS-AGFNLOGNGEAVDFVSTYDVTNFDAGNATTA-----	1300
DB	1152	ILGANGAFGQTSLLDITASGASANIINGYSQTGAVTNVCTVTLGSGGVLTSGLLTNGGILD	1211
QY	1301	-----KVYDDTSK-----TSKVYVDVNVDDTTIEVKDKLIG-----VKTTTLTSTG	1342
DB	1212	LTCGALNLTXXGGAFTVAGGLTGTAGTLNNGNLVSAAANSGLSGQTHIADVASVLTDTG	1271
QY	1343	T-----GANKFALSNOATGDALVKASDIVAHLNLTLSGDTQDTAKGASQANN	1387
DB	1272	TLCSTSAVEVLGTNLNAGANA-AMTNVLSGDTINTNAAV---TLSG-----NN	1315
QY	1388	S---AGVVDADGNKVYDSDTNKYQAKNDGTVDKTKVEAKDKLVAQAQTPDGTQAQMY	1444
DB	1316	SFSGAHOIGDGBELTV-----GQSNLGAASSATVNL-----GTLTSHLI	1354
QY	1445	KSVINKBQVDANKKOGINEDNAFVKGLEKAASDNKT-----KNAAVTVGDNLAV-	1494

```
Db 1355 LNVSESIANVLSGVAGTVD--IIGADTALTANNSGFLGOYALAGNSKLTVASTNNLG 1412
Qy 1495 AQTPLTFAAGTGTAKKLGELTIKG--GQTDNKLTDNNIGVVGAGTGGFTVKLAKDITNL 1553
Db 1413 ASSSVALLAG-----AGDTLSLSGFNCIFGNSVTGS--GVLOQVTDAAEV-----TLTSS 1458
Qy 1554 NSV--NAGGTKIDDKGVSEFVDSGQAKANTPVLSANGL-----DLGGKV----- 1595
Db 1459 NGVSNVYTIADIATLNDLIDIA-----LFNHVLTGNGLLNVAKNDASTAFDGFVGGAFFS 1514
Qy 1596 -ISNVGKGTDTDAANVQOLNEV-----RNLGLG--NAGNDNADGNOVN 1637
Db 1515 GIVNLWNTTFALSADNAALARAFLKLSDSVTTVGATDRTLHGLDLNGGTLIFDGS--- 1571
Qy 1638 IADIKKPNSSGSSNRVT---IKAGTV--LGGKNNDTKELAT-----GGIQVG 1681
Db 1572 -----PQSOAGVVTVDLALNSGTISITGACNWEHEPVPVPPNVSLLLBQDRGDILLE 1625
Qy 1682 VDKDGNANGDLSNVVKTQDKGSKALLATYNAAGQNTYLTNNPAEADIRNEOGIRPFH 1741
Db 1626 LINAANVTGNANNLDLLV--DGT----AITSQTGVESAIOQGGSTVANATHNYGLTSSN 1679
Qy 1742 VNDGNQEPVVOGRNGIDSSAGKHSVAIG-----FOAKADGEAAVAIGRQTQAGN 1791
Db 1680 GNGSGGLVYNTLSALELLANGANALLATESGLTANRVLNAELFGVGLVY--DAONGA 1737
Qy 1792 QSIAGDNAQATGQSIAGTGNVAGK---HSGAIGDPTVTRKADNSYSYVGNNNQETDA 1847
Db 1738 LTLANGNRYE-----GTTVTAGELILCANGAFQTSLLNASCASANINGY---- 1785
Qy 1848 TQTVDFVGNMIVTESNVALGNSAISAG--THAG-----TOAKKSDGTAGTTTTAG-- 1899
Db 1786 RQT-----VG-----AVTNSGAVTLNGGVLTSGLLTNGGILDLTGGALNLAAGGSGTVAGGL 1838
Qy 1900 -ATCTVAGFACQAVGAVSVGCASCAERRIONVAGEVSATSDAVNGSOLYKATQSIANA 1958
Db 1839 TGAGTLNINGDLAVSATNSGLSG--OHIADVASVTLTGTGLTSAVEVLGTL----- 1891
Qy 1959 TNELDHRHQENKANAGISSAMAMAMPOQAYIPGRSNVVTGGIATHNGQGAVALVGLSKLS 2018
Db 1892 -----NLGANAAMTNVLS-----GGVYINTNAAVTLCNNNSFS-GAHOIG 1931
Qy 2019 DNGQWFKINGSADTQGHVGAAGVAG 2044
Db 1932 TDGELTV---GQASNLGASSATVNLG 1954

RESULT 9
Q92D91 PRELIMINARY; PRT: 2340 AA.
AC Q92D91;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CELL SURFACE ANTIGEN (SCA3).
GN RP451.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsia;
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MADRID E;
RC MEDLINE=90039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Siceritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
DR EMBL: AJ235271; CAAL4908.1;
SQ SEQUENCE 2340 AA; 247862 MW; CC4070F93C165F26 CRC64;
```

```
Query Match 5.6%; Score 582; DB 2; Length 2340;
Best Local Similarity 21.2%; Pred. No. 3e-16;
Matches 499; Conservative 313; Mismatches 879; Indels 658; Gaps 119;

Qy 25 SHSTGGSCATQGVSVCTLSFARIAALAVLIGATLGSAY-AOKKTKHIAIGEONQ 83
Db 252 AYDLGSSSIITGDIIGNISLD-----TINVLLGSAFNSITLKATNINLKH----- 297
Qy 84 RRSCTAKADGDRATAGENANAGQAIAIGSSNKTVNGSSLDKIGTDATGQESIAIGD 143
Db 298 -NTSTLND-DNIIVIG---NIKNNNKDI--LNFKVHGTNLDNEMIIPAOKT---HGT 347
Qy 144 VKASGDASIAIGSDDLHL-----DQHGPNKPKG-TLINDLINGHAVLK-----EIRSSKD 195
Db 348 LNFKNATLNGNINNLNLFKFSGGHKTNLQNGTKVDNLVFAFSDVLSGTSVNGLLDT 407
Qy 196 D-VKYRRTTASGHASTAVGAMSYAOGHFSNAFGRATKASAYSLAVGLAATAEGOSTAI 254
Db 408 DCVTFNNSNVNG-GTLIINAKNTISAKLLNA--TKAKIQINANLTMN--HPSAGDISDRI 463
Qy 255 GSDATSSSLGAIALGAGTQALOGSI-ALGOGSVVVTQSDNNSRPAYTPNTQALDPKFOAT 313
Db 464 -ADNTIYTIDA-----KGNVNLLNNNAKIIFEGADSMALALINTGVTDRTFTIY 512
Qy 314 NN-----TKAGPLSIGSNSIKR--KIINYGAG---VNKTDA 344
Db 513 NNLQSGNDEYGVKIEAIKKVITIANOSGPTYIGQDNTHRLKELIVEGAGDIIDDIIF 572
Qy 345 VVNAQLEAVVKAERITFOGDDNSTDVKIGLDNLTJIKGAETNALTDNNGVNVKREAD 404
Db 573 TKLLSINSTGOITFRTLDLAGGN---IAFGKHGTLVNGVTGTSITSENNOGIL--TI 627
Qy 405 NSGLKVLAKLNLNLTTEVNTTLNATTVKVSSSTTAEILLSLSTTOPN----- 456
Db 628 NSG-----NITGVIGTNELGLKLVNIGADPVTCSANFASVALTNFSSVILADG 677
Qy 457 ---TGSQSTKTVGVNGVKFTNNAETAAIGTRITRDKIGFARDGVDDEKQAPYLOKK 513
Db 678 VTLTGVEVTHNTKGV--LSLGTGSGNITGQIGTNSAALEKI----- 716
Qy 514 QLKVGSAITTDNGIDAGNKKISNLAKSSANDAVTIQKAAKPTLNAGAGISVTPTEI 573
Db 717 --NIGAGASNIDSIYAGSTVLTQDTSELTNNDVVVNSNIITTAGNNSGK----- 765
Qy 574 SVDAKSNVTAPTNYNIGVKTTELNSDGTSDKFSVKGSGTNNSLVTAHLASVLYNVRFTA 633
Db 766 LIFTGNGGIGT---NIGANGAALQ-----EVVFNGTNN-----IGGTA 800
Qy 634 DSALQSTVTKBEDDDANAITVA-----KDT---TKNAGAVSILKLKGK----- 674
Db 801 NS--QNFTVAH--SAANVVTGLTTGALKYKDTGTIAHGLGVGDIDFNKAKGFIKGD 855
Qy 675 -----NGLJVATKK---DGTVTFGLSODSGLTGTCKSTLNNNDGLTVKDTNQGIVG 721
Db 856 GAMIDGSVLCNGGVAGTLDIFDGGNVVTQNIAGDANANSI--STINTQG---DNTKNVTI- 908
Qy 722 ANGKFTNVNGSNPGTGANTARTDKIGFAGSDGAVDTNKPDLQDKQLQGVNKIYNT 781
Db 909 ANDIFVDNIHFTNGI-LQLGNNLTHNHIDFGANGGTLEFN-----GN-----NT 952
Qy 782 GINAGGKAITGLSPTLPISIAQOSSRNIELGNTIQDKKSNAAASINDILMTG-----FN 834
Db 953 -YLNALIVNGONGNLNAFTNLKASDDTIGTV-----KIINIGIGTPQNET 998
Qy 835 LKNNNNPDIFFSTYD-IVDFANGNATTATVTHDHTANKTSKVYVDVN-----VDDTTI 885
Db 999 IQVNNKNTLVSSVNSSINF--GDANSQILSAPVDQTIKFNNLNELTGGGIITLDSNGN 1056
Qy 886 HLTGTDGDNK-KIGVKTTKLNTSANGN-TATN-----FNVNSS---DEDALVNAKD 931
Db 1057 NLTISGNGIKLGSKGNELSLNITKGVTVTNDLDTQNIHQNLNNGALFDDQSLTSAK- 1115
```

QY 932 IAEENLTAKETHTTKGADTALQTFYKVDENNADANAIVGQ-----KNAN 982  
 Db 1116 -IKNIN-----IGTVAGGA-----TYTDAINDFDLNTSGWFKHODSILELNSSNTN 1164  
 QY 983 NOVNTLTLAGENGLN-----IKTDKNVTTFGINTTSGL---KAGKSTLNDGG 1027  
 Db 1165 DHTITLTSALDPGNNOFGIILKLTOTNKLITIDNNGNAVYTLGTANHMLKQLTFASIDNGA 1224  
 QY 1028 LSKNPTSEQIQVADGVKFAKVNNGVVGAGIDGT--TRITRDEIGFTGTNGSLDKSKP 1086  
 Db 1225 IALKVGINVENTLNTKDJLELVANVLFNKNTTYTATNGINGHVDFQGNAGVI----- 1279  
 QY 1087 HLSKDGINAGGKKITIQSGEIAONSHDAVGTGKIYDLKTELENKI-----SSTAKTAONS 1142  
 Db 1280 NLNDDIEDGVSSTGVNGTFLNFGSGRVTG-----LINNIVMLQAGAGDVLSLA 1330  
 QY 1143 LHBFSVADEG-----NNPT--VSNPYSSYDTSKT-----SDV 1173  
 Db 1331 SGNYSITEIQGNGNNLTFANSHLTTDINKTGGQDLNLVFNINGSVSGSIGANAAGVDI 1390  
 QY 1174 ITFAG-----ENGITTKVNGVVRVGID--QTKGLTTPKL-----T 1207  
 Db 1391 IINAGSVNFNTLKSNGIVISDQATMOVNNVTATDISGNANNGTFLKNNHTPINITST 1450  
 QY 1208 VGNNGKGIVIDSONGQNTIT-----GLSNTLANVTNDKGSVKT 1246  
 Db 1451 LGNNAIG--TIEVANNDVTITGTLOAONIHFSNATQAATLTLAGAASQVNTITAGNNIHT 1509  
 QY 1247 TE-----QGN--ITKDEKTKRAASIVDLSAGFNLOGN-----EAVDFVSTYDVTNFPADG 1295  
 Db 1510 LEVTFDFTGNDGIIGDAN--NRLKSI-----ELTNGTIVTINSPHVYSITANNAGQ 1560  
 QY 1296 NA-----TTAKVTYDTSKTSKVYDVNVVDDTTI---EYKDKLGVKTTTLTSTGTGANKFAL 1350  
 Db 1561 NVKLNIEGGITVGLSKTSKLANVOISEDTTIRGDVYSKYLVNIDAGKTINFRDGNMNP 1620  
 QY 1351 SNAQTGDALV-----RASDIVAHLNTLSGD--IQTAGASOAN--NSAGYVDAD--GNKVIV 1401  
 Db 1621 KNLDPDALIDLVLPLSLFNTFTDIKADNLNFADDTATANEKDAVVIDAHIDNGGIL 1680  
 QY 1402 DSTONKY--YQAKNDGTVDKTEKAKDKLVAQAQTPDGTTLAQMVKV-----VI 1448  
 Db 1681 KFNDAWLTQEIKNANII-----EIASDKFM-----LLQKNIKAATLIADNANLVLL 1727  
 QY 1449 NKEQVNDANKKGINEDNAFVKGLEKAAASDNKTKNAAVTVGDLNAVAQTPLTFAGDTGTT 1508  
 Db 1728 DMVEVNTNLNVNRDIVLDIA---NYELKTYGNVTHNGLLTI-----ITY---FDTA 1771  
 QY 1509 AKKLGELTITKGGQDTRNKLTDNNIGVVVAGTGGFTVKLAKDLTNLNS-----VNA 1558  
 Db 1772 LQKGGHILVSQGSNVDMSDL--DNLI-----IKIKASHDITNITSDKHQIYKLETGA 1822  
 QY 1559 GGTKTDDKGVFVDSGQAKANTPVLNANGL-----DLGKGVISNVG-----KCTKDTDA 1609  
 Db 1823 IYTPVPQTVIIIDASEQNKFKVKKVADANGLVLLTDTGCR--DGTGGRDDTGRGNTD-- 1878  
 QY 1610 NVQOLNEVNLLGLGN-----AGNDNADGNQVNIADIKK--DPN-----S 1647  
 Db 1879 -----NCCRDNCVGNISNNSNEAGGSSDDKN--YGITDVVPFIDPSPILDYTKNNYVAS 1932  
 QY 1648 GSSS-----NRTVTKAGTVLGGKGNNDTEKATGQIOGVQDKG-----NANG-----DLSNV 1695  
 Db 1933 GIANQLINHVKFGNTDAGKLLNDLGFMSPNRVTTETDLRLSNRINVLNGLNEGVLNGI 1992  
 QY 1696 WVK-----TQKD--GSKKALLA---TYNAAGQNTYLTNNPAPADIRINEQGITR 1738  
 Db 1993 EVENFLTDIANDMNFETAKEIGNRLEELSDANTVNLKNTNLLNKK-----IN 2041  
 QY 1739 FFHVNDGNQBPVVOGRNGIDSSASGKHSVAIGFOAKADGEAAVATGROQTQAGNQSIAGD 1798  
 Db 2042 LKRLNTNQAIIAAGDE--DNIVTIGWMSFYGKIKONSKNS--ASGYQSNWCGGIIIGFDY 2098

QY 1799 NAOATGDSOIAIGTGNVAG---KHSIGAIDPSTVKADNSYSGV-----NNNQPTDA--- 1847  
 Db 2099 NI-----DNSIVIGAAYTMADSKVKHKNDKNGDRTRAKSNIYSIYGLYNWLTNNFEVEAIG 2154  
 QY 1848 -----TOTDVEGVGNITVTESNSVALGSNSAISAGTHAGT----- 1883  
 Db 2155 VYGRNKKIKNYEKRTITTDQIAIGKFINTFYSYELGGYNYLIS---HRTTITPFGMRY 2211  
 QY 1884 QAKSDGTAGTTTAGTGTGVKGFAG--QTAVGAVSVGASGAERRIQNVAAAGEVSATSTD 1941  
 Db 2212 ATFKNNGYKENNTTFQNLSTKKNYDKFTILGLNSV-----THYLSQD 2255  
 QY 1942 AVNGSQL-----YKATOSIANATNELDHRHONENKANAGISSAMAMASMPQAYIPGRSM 1996  
 Db 2256 IIKPELHWFYINQCKNLPNIDARLD-----GIDEPLTTIRFKPAKITYN--- 2301  
 QY 1997 VTGGIATHN 2005  
 Db 2302 LGGGISTKN 2310

RESULT 10  
 Q9ZHL0 PRELIMINARY; PRT; 4919 AA.  
 ID Q9ZHL0 AC Q9ZHL0;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
 DE LARGE SUPERNATANT PROTEIN 2.  
 GN LSPA2.  
 OS Haemophilus ducreyi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=730;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=35000.  
 RX MEDLINE=99030326; PubMed=9811662;  
 RA Ward C.K., Lumley S.R., Latimer J.L., Cope L.D., Hansen E.J.;  
 RT "Haemophilus ducreyi secretes a filamentous hemagglutinin-like  
 protein";  
 RL J. Bacteriol. 180:6013-6022(1998).  
 DR EMBL: AF057696; AAC79761.1;  
 SQ SEQUENCE 4919 AA; 542602 MW; 5779201455CA69A0 CRC64;

Query Match 5.5%; Score 570.5; DB 2; Length 4919;  
 Best Local Similarity 19.2%; Pred. No. 2.5e-15;  
 Matches 484; Conservative 353; Mismatches 914; Indels 771; Gaps 109;

QY 2 NHTYKVFKNKATGTFMAVAYAKSHSTGGSCATQGVGS-----VCT----- 43  
 Db 3 NKRYKLIFFSKVKNCVPAENIKSASGSGSSNSKIAEDQEEEDPSLACSLSSSIH 62  
 QY 44 -----LSTFARIALAVLVIGATLSCSAVAQKKDKTKHIAIGBONOPRRS- 86  
 Db 63 LGLRHNSPLKVFVKGNLVSLLSLMPAMQVWADSSNAIVDHSIGAKQTAVDERD-PKNGK 121  
 QY 87 -----GTAKADGDRAI-----AIGENANAQ--GOAIAIGSNKTVNGSSLD 126  
 Db 122 EKVVIINIAKPD--EQLSDNHFSKFNLPNSAVFNNSIKESNSQLVGLLGENKLGSAQA 180  
 QY 127 KIGTDATGQESIAIGGVKASGASSTAIGSDDLHLLDQHGHNPKHPKGTINDL--INGHA 184  
 Db 181 TIFNQVTGDOESKISGGLEVFEKA-----DLFII-----NPNGVTLNGVKTINTDR 227  
 QY 185 VLKEIRSKNDVK-----YRRTTASGHASTAVGAMSYAOGHFSNAFCTRATKAYSALV 240  
 Db 228 FVASTSEVPEPHIKOLNVQRKVIIGKGVATNGLS-----HFD----- 266  
 QY 241 GLAATAEQSTIATGSDATSSSLGAIAGTAGTAAQLOGSIALCGSVVTSODNNSPAYT 300  
 Db 267 VVAKNIEQGVKSTEGSKPAKLANVTFAAAGNLTY-----DVTNRDVT-N 309

Qy 301 PNTQALDPKFOATNNTKAGPLSGSNS-----IKRKINVGAGVN-----KTDVAVN 346  
D 310 RNT---NPKRPITDNTKONIAISGESAGSMYGRNIKFIPTVDKAGVNHQGVIFAEDDIN 366  
Qy 347 V-----AQLEAVYWKAKERRITFGQDDNSTDKVIGLDNLTITKGAETNALT 393  
D 367 ILTDGNSRLNKVYADVVRV--GKOTELANNGQIHA--DOOLIILNATGHVKLNDGSSVIS 423  
Qy 394 DNNIGV-----YKEADNSGLKYKLAKTLNNLFEVNTTTLNATTTVKVGSSTTA 443  
D 424 NNLGISALNLTLENATVSANLSFRVNTDKLNLKSVSARAAD-----LQSGNLNDKA 479  
Qy 444 ELLSDSLTFPTQPTGSGOSTSKTYGVGVNGVF-----TNAETTAAGITRITRDIGF 496  
D 480 SVLAHKLTILNISDVSLN--NOSKLSANLKIKKVRLDNLNANSELSE-----NNLT 529  
Qy 497 ARDGDVDEKQAPYLDKKOLKVGVAITDNGID-----AGKKISNLAKGSSANDAVT 549  
D 530 NTSNNTLK-----NKSKEFTAGNMTLVNTVNTLNNDSELAANLTLNVTKNVTLNDA-- 582  
Qy 550 IEOLKAAKPTLNAGAGISVTPTEISVDAKSGNVTAPTINIG-----VKTTELSNDSGTS 603  
D 583 -SKLSANKLDLN-----VTDNVTLSKS-----TLSAGELTFKKVKVNTLNDSE-- 626  
Qy 604 KFSVKGSGTNSLVTAEHLASYLEVN---RTADSALQSFVKEEDDDANAITVAKDTT 660  
D 627 -----LAANNLSLASHNVTLNKSKLSAQKADIKAVNLTLNDTTELAKNLDINSTTI 680  
Qy 661 KNAGAVSIL-----KLKKG-----NGLTVATKKDGTVTGLS 692  
D 681 TNGTIAGIFANITTEKLNKKEKALILAEONLNTVNGSHYENKGDIVSKDATTESKN 740  
Qy 693 QDSGLTICKSLNNDGLTVTDNQIOVGANGIKFTNVNGSNPGTGIANTRIDRIGF 752  
D 741 SD-----FTSNGSKLVNAQOLKAVNNFTISO-----GDDITLIGNVTLNASGT 785  
Qy 753 AGSGAVDTNKPYPDQKLVGNV--TNTGINAGGKAITGLSPTLPSIADQSSRNIELG 811  
D 786 FTNSGNLTUVK-----TLDVGDIQNFNGK-----NLTVG 815  
Qy 812 NTIQDKKS-----NAASINDLNTGF-----NLKNNNPIDF 844  
D 816 EDLHKSKTKINDGKLISLKNLISSEADFINGTLLGTEALKIATKGNFTNEKAILA 875  
Qy 845 VSTYDVIDFANGNATTATVTHDTA---NKTSKVYDVNDVDTTH-----LTGTDN 893  
D 876 SNLSDISVAEGKTFNNGTIESGKNLNTNTGAF-LNVDNATIRSEGLNITSTGVSN 934  
Qy 894 KKLGVTKTKLNKTSANGNTATNFVNSDDEDALYNKADIAENLNLAKEIHTTKGTADTA 953  
D 935 NGTLISNERLNTS-----AANF--TNEISNGTVMISNGL-----LNIAKQGNITNKNLIAS 983  
Qy 954 LOTFTVKKVVDEN--NNADDANAITVGOK-----NANNQVNTLTKGE---NGLNKTD 1001  
D 984 RQQLNLTAVDNITDNTNISNIAVLISLGNLSNKSQVYNL---GEIYAGNNISVAH 1040  
Qy 1002 ---KNGTFTGINTTSGLKACKAGTSLNDGCLSIKNPTGSEQIOVGADGVKFAKVNNGVYGA 1059  
D 1041 QLKNDVKLMGDITK-----TKGOASYKLYQASNGHFGNDGSSGYSEGDLNKGK 1092  
Qy 1060 GIDGTTTRITREIG--FTGTNGSLDKSKPHLSKOGINAGK-----KIT--- 1101  
D 1093 FADLONKLTQIRIGKIVAGROLTENKS-----NAGGKSEIINRGITNVNKNLSYDS 1143  
Qy 1102 -----NIQSGEIAQNSHDAVTGKIKYDLKTELENKISSAKTQANSIHEFSVADEQGN 1155  
D 1144 DVSENNMQSOKV-----DLYT--KIFEAKSDIE-----LTFKT--NGTHPVYL-----N 1184  
Qy 1156 FTVSNPXYSSYDTSKTSOVIITPAGE---NGIPTKVNKGVVVRGIDQTKGLTTP---KLTVG 1209  
D 1185 FKSNNKERYNSENTRNFKSIGDLNLEALSAPAEIAEYVSGSSSNYINPVSYLAALG 1244

Qy 1210 NNNGKGIVIDSQONQNTITGLSNTLAN-----VTNDKGSVRTTE 1248  
D 1245 NAN-----NSSNPHYLNTALKHILGNQDDLLKKQENIKVLKOKKEDFKDKGASKMLD 1298  
Qy 1249 QGNIIKDEDKTPRAASIVDVLSSAGNLOGNEBPAVDFSVTYDVTNFPADG---NATTAKVTYD 1305  
D 1299 ---LYPNTDEKAKIFAGIIRNGNDTISDVESEDFKKYS--KFQNGEWAKNDTGTDSDYD 1353  
Qy 1306 DTSKTSVVYDVNVDDTTIEYKDKKLGVKTTTLTSTGTGANKFALSNOATGDALYKASDI 1365  
D 1354 STKASEYKVENVDHKE-NIDEHKLNIKGGHEITVPGVSFEN--LNNKNMHDQPKLGEI 1410  
Qy 1366 VAHL--WTLSDGIOTAKGASQ-----ANNSAGVVDADGNKVIYDSTDN---KYIOAK 1412  
D 1411 DKSTIISLEAOPVYTEKSAARDSPRVQNQDKEALDNLRYTRLSYINQNNYLKAFYFNQ 1470  
Qy 1413 NDTGVDTKEVAKDKLVAOAQTPDGTLAQMNKSVINKVEQVNDANKKOCINEDNAFVKGL 1472  
D 1471 LDTEDDKLKGIKR-----IGDNYFEHOLITRLIEKADVADNHLTKHGLH-DIALVKKL 1521  
Qy 1473 EKAASDNKTKNAAVTVGDLNVAOQTPITFAGDTGTTAKKLGELT----- 1517  
D 1522 IDSAS-----IQAKDLNL-----KVGELTKKQKDLNKEDIVWV 1556  
Qy 1518 -----IKGQTDNTKLTNNNGVWAG-TDGF 1542  
D 1557 KTEVNGEVLVPQVYLAKQTIEVEKORGVGQTQIRAGIIDVKVDDVNTGTIAGYAVGL 1616  
Qy 1543 TVK-----LAKDLNLNSVNAAGTKIDDKGVSFVDSG----- 1575  
D 1617 EAKNKLKNTGILSQRLSK-----VGKKGLESTGVTVYVDETGATKVKRKARIKSEGHILE 1672  
Qy 1576 -----QAKANTPVLSANGLDLG-----GKVISNVKGFKTDTA 1608  
D 1673 TDKDNVDLTASELKGNTQIKAKDLNLDIYETSYKYKELFKCKNGGEIGDRVQTQSQ 1732  
Qy 1609 ANVOQLNEVRNLLGLNAGNDNAGNQV---NTADIKKDPNSGSSNR-----TVIRAG 1659  
D 1733 AKSVGTDAFPHLHLSLEGVDNQTGSNLKANRTTGVVRGDFNTRKAGDLFHRQIDVTSG 1792  
Qy 1660 TV----LGGKNNDEKLTATGGIOGVDBK---DGNANGDLSNVWVKTK----- 1701  
D 1793 TVYSASASGGQSAGISLTDQGVETYNKTATAGNADVTNFMKRTETETSLTHRNSEF 1852  
Qy 1702 -----DCSKKALL-----ATYNAAGQ 1717  
D 1853 NALSGELVVMGKADIGGVDINRDVEV IKTPEIAAEQKAAEBAKAEVKEASEATAAKE 1912  
Qy 1718 TNYLTNNPAEAI DR-----INEQGIREFHVNDGNQBPVVOGRNGIDSSA 1761  
D 1913 TEEABNDNVABEKDKTKPKPKKLTDDEIAAFAFETKGEDEFAAYKAREEE--DRKKGFTLSA 1970  
Qy 1762 SGKHS-----VAIGFOAKADGEAAVAIGRO-----TOAG---NOSIAG 1797  
D 1971 EQIESTKARDEKETYYIELKVGVGAEAAHSAADAIENKARQIIDTQNLKQDQTVLQ 2030  
Qy 1798 DNAQ-----ATGDQSIATIGTNVAKHSGAICDPSTVK-----ADNSYSVGNNOFTDAT 1848  
D 2031 EASDVNLATGD-----LAGASAKLKFELSTIEBKSRGSDGRSILGRLNLAAR 2080  
Qy 1849 QTDVFGVGNNTVTESNVALGNSAISAGTHAGTQAKKSDGTACTTTTAGATGVTKGFA 1908  
D 2081 GGDI--TLNNYETTESHLKARDNVNNSGV-TEQDESNQSLKVTAGASSCGGVA 2137  
Qy 1909 GOTAVGAVSVCASGAERRIQNVAAGEVSATSTDAVNGSOL-YKATQ--SIANATNELDHR 1965  
D 2138 GGCSAG-VSAGVSGS-----YNESNTFESTSHTNLSLRGSLRVEAGKDFNLISNNVDVH- 2191  
Qy 1966 IHQENKANAGISSAMAMAPQAYIPGRSMVWTGGIATHNG---OCAVAVGLSKLSDNQ 2022  
D 2192 LHL-DVKGDTNVVSKQDSYSRKERGNYSVSAGVGVSTAGGARPNGSVGLGVSAENENSK 2250  
Qy 2023 WV 2024



```
Db 1335 I-----VLGASNLNRVNGTAMLVKVGAGNVTIAKGGNVKIGEIOGTGTNT-----LTLPAH 1384
Qy 1712 YNAAGOTNYLTNPAAEADIRNEQIRFFHVNDQNPVQVQGRNGIDSSASKHSAITGF 1771
Db 1385 FKLTC-----SINKTGGQALKLNFWNGSGVGV----- 1413
Qy 1772 QAKADGEAANAIGROTQNGNSIALGDNAAQT-----GDQSTIA-----IGTGNVVAGKHSGAI 1824
Db 1414 -----GTAANSVGDITTAGATSFASSVNAKGTATLGGTTSFANTFTNTGAVTLAK----- 1463
Qy 1825 GDPSTVRADNSYSYGNNNQFTDATQTFVGVGNNTVTESNSVALGSNSALSAGTHAGTQ 1884
Db 1464 -----GSITSFAKNVTATSF-VANSATINFGNSLAFNSNITGS----- 1500
Qy 1885 AKKSDGTAGTTTTAGA-----TGTVK-----GFAGQTAAGVAVSGASGAEERRIONV 1930
Db 1501 -----GTTTLCAQNVTVTGTGSGFTDTLTNTTFDGAAKSGNLIKSGSTLDLSGV 1552
Qy 1931 A--AGEVSNSTSDAVNGSOLYKAT 1952
Db 1553 SNLALVVTATNFMNINISPDTKYT 1576

RESULT 12
Q9XC47 PRELIMINARY; PRT; 2106 AA.
AC Q9XC47;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE OUTER MEMBRANE PROTEIN A.
GN OMPA.
OS Rickettsia australis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OC NCBI_TaxID=787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PHS;
RX MEDLINE=20487299; PubMed=11034486;
RA Stenos J., Walker D.H.;
RT "The rickettsial outer-membrane protein A and B genes of Rickettsia
RT australis, the most divergent rickettsia of the spotted fever group.";
RL Int. J. Syst. Evol. Microbiol. 50:1775-1779(2000).
DR EMBL; AF149108; AAD39531.2; -.
SQ SEQUENCE 2106 AA; 209075 MW; A12B28F54BEEDECC CRC64;

Query Match 5.48; Score 555; DB 2; Length 2106;
Best Local Similarity 21.78; Pred. No. 3.5e-15;
Matches 506; Conservative 239; Mismatches 782; Indels 806; Gaps 117;

Qy 1 MNHIYKVFENKATGTFMAVAEYAKS-----HSTGGGSCATGQVGSVCTLSFARIAAL-- 52
Db 1 MANISPKLFQRAIHKSLKAALFTSTAAILSSSGAWGVAAGVI-SVNDAAFSNRAVANN 59
Qy 53 -AVLVIGATLSGAYAQKDKPKHIAIGQNPQRRSGTAKADGDRAIATGENANAQGGQAI 111
Db 60 WNEITAGGAANGNADGPDQNEAFTYG-----GNHTITADEAGRIITAINVAGTTPV 111
Qy 112 AIGSSNKT-----VNGSSLDKIGTDA-----TGQESIAIG 141
Db 112 ALNSTQNTSVGSIIVTGGNLLPVTIADGKSLTLTGTKVAADHGFDAADNNTGLGAIAG 171
Qy 142 G-----DVKASGDASI-----AIGSDLLHLBQHGPKHPKGTTLNDLINGHAVLKEIRSS 192
Db 172 GATAGLTITQSATPALINLAGAIDGN-----NDHGEITVNTTSTGTGIGGKTLHAVTFN 227
Qy 193 KDNVYKRRRTTASGHASTAVGAMSVAOQHFNAGFTRATAKSAYSILAVGLAATAEGQSTI 252
Db 228 NGN-----ACGVSTA-CAIS-----ATTV 245
Qy 253 AIGSDA-----TSSSLGAIAGTAGTQAQGLQSGIALGQGSVVVTQSDNNSRPAYTPNQ 304
```

```
Db 246 MIGEDAGNVAHVQVOTGLITGAVNFAADGALTANNGIA---GAVTT----- 287
Qy 305 ALDPKFOATNNTKAGPLSIGS-----NSIKRKIINVAGVKNKTDVNV----- 347
Db 288 -----ANNNT--GTLTVGAGDVTGAIGVNGGNILKQVLFNGASNVATIDATNTVTTINNA 338
Qy 348 -----AQLBAVVKAKERRITFGD-----DNSTDVKIGLDNTLTITKGAETNAL 392
Db 339 AANVTAGAITAAVNAFADGALTANNGLAGAVTTANNNT-----GTLTVGAGDVTGAI 391
Qy 393 TDNNIGVVKEADNSGLKVKLAKTLNNLTVNTTTLNAT--TTVKVGSSTTAELLSDSL 451
Db 392 GVNGGNILKQVLFNGAS-----NVATIDATNTVTTINNAANVTAAAGAITAAVN 438
Qy 452 FTQPNMTGSQSTSKTVYGVNGVKFTNNAETAAIGTRITRITRIGKIFARDGDVDEKQAPYLD 511
Db 439 FAA--DGALTANNGLAG--AVTTANNNTGTLTVG-----GDV----- 472
Qy 512 KKQLKVGVAI-----TIDNG-----IDAGNKKISNLAAGSSANDAVTIEQLKAAK 557
Db 473 -----TGAIGVNGGNILKQVLFNGASNVATIDATNTVTTINNAANVTAAAGITAAVNAFAD 527
Qy 558 PTLNAGAGISVTPTEISYDAKSGNVTAPTNYNIGVKVTTIELNSDGTSDKFSVKSGTNNLSV 617
Db 528 GALTANNNGI-----AGAVTTANNNTGTLTV-----GAGDVTGAIGVNGGNILK 570
Qy 618 T-----AEHLASYLNEVNTADSALQSFVKBEEDDDANAITVAKD-----TTKNAGAV 666
Db 571 QVLFNGASNVAT-IDATNTVTTINNAANVT-----AGAITAAVNAFADGALTANNGLV 622
Qy 667 SILKLGKNGLTVAATKDGTVTFGLSQDSGLTIGKSTLNNDLGLTVKDNEQIQVGANG-- 724
Db 623 -----GAVTTANNNTGTLTVGAGDVTG-AIGV---NGGNILK-----QVLFNGAS 663
Qy 725 ----IKFTNVNGSNPCTGIANTRITRDKIGFAGSDGAVDTNPKPYLDQKLVGVNVI--- 778
Db 664 NVATIDATNTVTTINNAANVTAAAGITA-AVNFA-ADGALTANNNGI-----VGAVTTAN 714
Qy 779 TNTG-INAGGKRAITGLSPTLPSIADQSSRNIBELGNTIOD---KDKSNAASINDILNTGFN 834
Db 715 NNTGTLTVGAGDVTG-----AIGVNGGNILKQVLFNGASNVATID-----ATN 757
Qy 835 LKNNNPIDFVSTYDI---VDFANGNATATVTHDTANKTSKVYVDVNVDDTTHLGTGD 891
Db 758 VTINNAANVTAAAGITA-AVNFAADGALTAN-----NGIAGAVTTANNNTGTLTVGAGD 811
Qy 892 DNKKLGKVTTLKNTKSANGNTATNFNVSSDEDAVLNKAIDIAENLTLAKEIHTTKGTAD 951
Db 812 VTGAIGVN-----GNILKQVLFNGASNVATIDATNTVTTINNAANVTAAAGITAAVNAFAD 858
Qy 952 TALQFTTVKVDENNADDA---NAITVGQKNANNQVNTLTLC-----GENGLNFK 999
Db 859 TAAVNFA-----ADGALTANNNGI-VGAVTTANNNTGTLTVGAGDVTGAIGVNGGNIL 909
Qy 1000 TDKNGTFTVF-GINTTSGLKAGKSTLNDGSLSTKNPTG-SEQIQVGADGVKFAKVNNGV 1057
Db 910 KQ-----VLFNGASNVATIDATNTVTTINNAANVTAAAGITAAVNAFADG---ALTANNGIT 962
Qy 1058 G-----AGIDGTRITRDEIGFTGTNGSLDKSKPHLSKDGINAGGKKITNIQSGETAQNSH 1113
Db 963 GAVTAAGANGT--LFTAGVGTGAVGT-----NAASLKVLN---AKYDNGAA 1004
Qy 1114 DAVTGCKLYDLKTELENKISSATAQNSLHSEFVADQEGNNFTVSNPYSDTSKTSDV 1173
Db 1005 DLVFTSDIY-----ARTV-----NFD-----DIAGGAV 1028
Qy 1174 ITF-AGENGITTKVNGVVRVIGIDQTKLTPKLVGNNNGKIVIDS--QNGQNTITGL 1230
Db 1029 GTIQVGNGLIATNVFG-----GNAQGT---LXLNGPVGKSYTLSTGLIANGNNATINI 1079
Qy 1231 SNTLANVTNDKGSVTTTEQGNLIKDEKTRAAAI-----VDVLSAGFNLOGNCEAVDFVST 1286
```



Db	1080	-NTLGLVTABNAXIGTVAQINI--QDNKTFANVKNADIEIL-----NAQAIDFKGA	1128
Qy	1287	YDTVNFADGNATTAKTVDYDTSKTSKVYDVNVDDTTIEVKDKKLGKVTTLTSTGTGCA	1346
Db	1129	NSKLFVNNSAT-----DDRVVTIKN-----DLPAPATGGG	1159
Qy	1347	KFALSNGATGDALVKASDVAIHNLTSGLDIQTAKGASQONNSAGVYDADGN---KVIYDS	1403
Db	1160	MLLFG-----TANLITLQGD-----GAXKAIGTAGNKLASNLVLGKVAFNN	1201
Qy	1404	TDNKYVQAKNDGVKTEKAVAKDLVAQAQTPDGTLAQMNKSVINKQVNDANKKOGIN	1463
Db	1202	IDTNNVLAFN-----IINTQFVD-----	1220
Qy	1464	EDNAFVKGLEKAASDNKNTAAVTVGDLNAVAQTPLTFAGDTGTAKKLGCEFTLTGQGT	1523
Db	1221	VGITNQIS-----VINIGAANA-----SLGLPAAAGSYTIDANG--	1256
Qy	1524	DTNKLTDNNIGVAGTDGFTVKLAKDLTNLNSVN--AGCTKI-----DDKGVSEV	1571
Db	1257	NVGILA--NGOTINFAHEDAEVLVIONSAGNSTITLNAALDPGAPNGKLA	1306
Qy	1572	DSSGOAKNTPVLSANGLDLGG-----KVISNVKGT--KDTDA-ANVOOL-----	1614
Db	1307	D-SCXAGRVILASVRNATYGTAVNKLKLEFRGNGTFOQIDTDFANLLELPVPAITYN	1365
Qy	1615	NEVRNLGLGNA-----GNDNAD--GNQVNIADIKKDPNSGSSNRTVIRAGTVLGGK	1665
Db	1366	KDINKNLSFGAATALTQNGINGVDFNQAAVITL-----GANKNIT-----GSVTSSN	1415
Qy	1666	GNNDTEKATGGOVGVKDGKNGANGDLSNVVVK-----NGHITNAMLKVGAGAVSITQGGTSTITEQNGTAL	1466
Db	1416	GVNQT-TIATGASTI-----NGHITNAMLKVGAGAVSITQGGTSTITEQNGTAL	1466
Qy	1709	L---ATYNAAGOTWYLTNNPAEADIRNEQGIREFHVNDQNPVQVGRNGCIDSSAGKH	1765
Db	1467	LTPANFNLTG-----SINKTGGQALKNFETNGSGVRGW-----	1501
Qy	1766	SVAGFOAKAGEAAVATGROTOAGNOSIAIGDNAQAT-----GDQSTA---IGTGNVAG	1818
Db	1502	GTANSVGDITTYTGTTFNFASSVNAKGAATLCGTTSPADFTNTGAVTLA	1550
Qy	1819	KHSGAIGDPSTVKADNSYSGVNNQFDTATQDTDFGVGNNTITVETESVAL-----GSNS	1873
Db	1551	KAS-----ITNFAKNVTATSF--AANNATINFGNSLAFNSNITGSGT	1590
Qy	1874	ALSAG-----THAGT-----QAKKSDG-----TAGTTTTAGATGT-----	1903
Db	1591	TTLTGASQVITYTGTGTFDTLTLNTFTDCAKSDGHILIKSGSTLDLGSVSTLALVVTAT	1650
Qy	1904	VRKGFAGQTAAGVAVSGASGAERRIQNVAAGEVSATSTD-----AVNGSOLYKATQ	1953
Db	1651	NFDINNISPNTKYTVISAETAGG---LKPTPAGNVKVTYVNDNRVNFVTFDESLTLFAE	1707
Qy	1954	STANATNELDHR-----IHONENKANAGISSAMAMASMPQAVIPGRSMV-----	1997
Db	1708	DIABEEVIEEDFKPGGPLANIPAAV---KKSLEMEQAPIGSDACIAFDWCKNLPIQ	1763
Qy	1998	TGCIATHNQGVAVAGLSKLSLSDQWVFK-----INGSADTQGHVGAAGVAG	2044
Db	1764	AAVATHLMQDVVKPSPDTIAANNQVILNSISSLINLNMARMDAIOQAAVAA	1816
RESULT	13		
Q9VTK8		PRELIMINARY; PRT; 2586 AA.	
AC	Q9VTK8		
DT	01-MAY-2000	(TrEMBLrel. 13, Created)	
DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)	
DE	01-MAR-2001	(TrEMBLrel. 16, Last annotation update)	
GN	CG18331	PROTEIN.	
OS	Drosophila melanogaster (Fruit fly).		

Db 279 TSSSDPVVEVAQGS--SSNGDGNSTQSSSTTTTTTTTTSSDGGQSTTSLSDPVVEVSQGTNG 337  
Qy 319 GPLSTGNSIKRKIIINGAGVKNKTDVAVNAQLEAVVWAKERRITFQGGDNSTDKVIGLD 378  
Db 338 GNSSTQSSSTTTTTSSDEGQTTSSDPVGEV-----AQGSSNGDGNSTQSSSTTTT 389  
Qy 379 NTLIKGAETNALTDNNIGVKEADNSGLKVLAKTLNLTENTVNTPLNATT----- 432  
Db 390 TTTTSSDGGQSTTSSDPVVEVSQGTNG-----NSSTQSSATTTTTSSDEGQTT 440  
Qy 433 -----VKVGSST-----STTAELLSDLFTQPTNGSQSTS-----KTVYGVNGVKF 474  
Db 441 SSSDPVVEVAQGSSSIGDGNSTQSTTTTTTTTTSSDGGQSTTSSDPVVEVSQGTNG--- 497  
Qy 475 TNNAAETAAICTRITRDKIGFARDGVDVEKOAPYLDKKQLKVGSAITIDNGIDAGNKK 534  
Db 498 -GNSSTQSSSTTTTTSS-----DEGQT-----TSSSDP 525  
Qy 535 ISNLAKGSAN--DAVTIEOLKAAPLNLNAGAGISVTPTEISVDKASGNVTAPTYNLGV 591  
Db 526 VVEVAQGSSSNGDGNSTQSSSTTTTTTTTTSSDGGQSTTSSDPVVEASQGT-----NGGN 579  
Qy 592 KTEILNSDCTSDKFSVKSGSTNNSLVTAEE--HLASYLNEVNKRTADSAQSFTVKEEDDD 649  
Db 580 SSTQSSSTTTTTSSDEGQTTSSDPVVEVAQGSSSIGDGNSTQSSSTTTTTTTTTSSDGG 639  
Qy 650 ANAITVAKDTTKNACAVSILKLGKNGILVATKKDGTVFLGSDQSLGTLCKSLNNDGL 709  
Db 640 QS--TTSSDPVVEAS-----QGTNGNSSTQSSSTTTTTTTTTSSDGGQTTSSS 683  
Qy 710 TVKDTNEIQVANGCIKFTNVNGSNPGTGIANTRITRDKIGFAGSDGAVDT--NKPYLD 767  
Db 684 ---DPVVEVAQGSST-----NGDGNSTQSSSTTTTTTTTT--TSSDGGQSTTSLSDPVVE 730  
Qy 768 -----QDKLVQGNVKITWGINAGKALITGLSPTLPSLTDOSRRNIELGNTIQQDK 819  
Db 731 CSQGTNGNSSTQSSSTTTTTSSDEG--PTTSSDPVVEVAQGSSSNGD--GNSQSTT 787  
Qy 820 SNAASINDILNTGNLKNNNPIDFVS---TYDIVDFANGNATTATVPHDTANKTSK--V 874  
Db 788 TTTTSTTS--SDGGSFTSSDPVVEVSQGTNGDNSSTQSSSTTTTTTSSDEGQTTSSSDP 845  
Qy 875 YVDV-----NVDDTIIHLGTDDNKKLGVTTKLNKTSANGNTATNFVNSSD 922  
Db 846 VSEVAQGSSTGDNSTQSSSTTTTTTTDIFRCPRQRTTTTTTTTTSSDGGESTTL----SD 901  
Qy 923 EDALVNAKDIAENLTLAKEIHLTKGTADTALQFTVKKVDENNADANALITVGOKNAN 982  
Db 902 PVVEVSQGTNGNSSTQSSSTTTTTSSDEGQTTSSSDPVVEVAQGSSTN-----GDGNS 957  
Qy 983 NOVNTLTL-----KGENGLNIKTDKNGVTFTGINTTSG----- 1015  
Db 958 QSSSTTTTTTTTTSSDGBESTTSSDPVVEVSQGTNGDNSSTQSSSTTTTTSSDGGQATSS 1017  
Qy 1016 -----LKAGSTLNDGGLSKNPTGSRQIQVAGDGVKFAKVNNGV-VGAGIDG----- 1063  
Db 1018 SAPVVDISQGSSSNGDGNSTQSSSTTTTTTTTTSSDGGQSTTSSDPVVEVSQGTNGNSST 1077  
Qy 1064 -----TTRITRDEIGFTGNGSLDKS-----KPHLSKD 1091  
Db 1078 QSSSTTTTTSSDEGQTTSSFGSCRSRSHRILFQCPROQLOLRHLPMVNPPLHLT-- 1135  
Qy 1092 GINAGGKXITNIQSGEIAQNSHDVTCGKIYDLKTELENKISSTAKTAQNSLHFSVADE 1151  
Db 1136 -----PVVEVSQGTNGDNSSTQSSSTTTTTSSDGGQTTSSAPVSSSTTTTTTTTTSS 1188  
Qy 1152 QGNFTVS-----NPSYSDTSKTSVDVITFAGENGILITKVKGVVVRVIDQTKG 1200  
Db 1189 DGDQSTTSSDPVVEVSQGTNGDNSSTQSSSTTTTTTTTTSSDGGQTTSSAPVVEV-----TQ 1244  
Qy 1201 LTTPKLTVCNNNGKIVIDSGONTITGLSNTLANVNDKGSVRTTEQGNLIK----- 1254  
Db 1245 -----SSSNG-----DGNSTQSSSTTTTTTTTTSSDGGESTTSSDPVVEVSQGTN 1289

Qy 1255 -DEKTRAASIVDVLSAGFNLOQNGEAVDFVS--TYDTVNFADGNATTAKVTVYDDTSKTSK 1312  
Db 1289 GDNSTQSSSTTTTTSSDEGQTTSSAPVVEVTQSSSNGDGNSTQSSSTTTTTTTTTSS 1348  
Qy 1313 VVYD--VNVDDTTIEVKDKLGVKTTTLTSTGTGANKFALSNOATGD-----ALVKASD 1364  
Db 1349 DGGESTTSSDPVVEVSQGTNGDNSSTQSSSTTTTTSSDGGQTTSSSGPVVDISOGSSSN 1408  
Qy 1365 IVAHLNTLSGDT-----OTAKGASQANNS 1388  
Db 1409 VLDNNDNNYDDIFRWANPPVLTPLWKSVEQMAITAPLSLRRPPQOQTSSDEGQTTSS 1468  
Qy 1389 AGYV-----DADGNKVIYDSTDNKYYQAKNDGTVDKTEKVAKDLVAAQAQPDG-- 1437  
Db 1469 SAPVVDISQGSSSNGDGNSTQSSSTTTTTTTTTSSDGGESTT---LSDPVVEVSQGTNGDN 1525  
Qy 1438 --TLAQMNVKSVINKEOVNDANKKOGINEDNAFVGLKAAKSDNKTNAAVTV----- 1488  
Db 1526 NSTOSSSTTTTTSSDGGQTTSSSDPVVE---VAQGSSSNGDGNSTQSTTTTTTTTTSS 1582  
Qy 1489 -GDLNVAQTLPLTFA-----GDTGTTAKKLGETLTIKGGOTDTNKLTDNNIGVAGT--- 1539  
Db 1583 DGGESTTSSDPVVEVSQGTNGDNSSTQSSSTTTTTSSDGGQTTSSSDPVVEVAQGSSTN 1642  
Qy 1540 -DGFTVKLAKDLTLNLSNVNAG--TKIDDKGVSVFVDSGG----- 1575  
Db 1643 GDGNSTQSSSTTTTTTTTTSSDGGESTTSSDPVVELRDLPLVMVMETLPSRPROQLQRHL 1702  
Qy 1576 -----QAKANTPVLS-----ANGLDL 1591  
Db 1703 PMVNPPLHLPLMKSVEQTAITAPLSLRRPPQOHLQMRALPLLSDPVVEVAQGSST 1762  
Qy 1592 GG-----KVISNVGKGTDDAANVQOLNEVRNLLGLGNAGNDNADG--- 1633  
Db 1763 NGDGNSTOSSSTTTTTTTTTSSDGGESTTSSDP--VVEVSQGTN-----GDNSSQSS 1812  
Qy 1634 -----NOVNTADIK-----KDPNSGSSSNRTVIKAGTVLGGKGNNDTE 1671  
Db 1813 SSTTTTNEVSLKDNRPKNWRTTKYSSRTIRIPNSGRKLSNSSSETSTTV---TSSSS 1869  
Qy 1672 KLATGGIQGVGDKGNANGDLNSVWVKQDKGSKKALLATYNAAGQTNVLTNNPAAIDR 1731  
Db 1870 KPQTKYSWSSSSKSNNGGKKNKYWKRTKKSRKNNNGSSSTIVGE-----ESSDS 1920  
Qy 1732 INEQGIRFFHVNDGNOE-----PVVQGRNGIDSSASG-----KHSVAIGF 1771  
Db 1921 LTDAGVDVTOGNGLNDGESSOSTVTSLLPVVDTSADYQNSSESLSTSTNTTKYSKSEK 1980  
Qy 1772 QAKADGEAAVAGROTQAGNOSIAIGDNAQAATGDSIAIGTGNVVGKHSIGAIDPSTVK 1831  
Db 1981 VPKSNQSSISASKTKTKVTSTSTSPNVKSSKK-----TSN--SGK---SVKTSSTFI 2030  
Qy 1832 ADNSYSVGNNOFTDA--TQTDVFGVGNNTIVTESNSVALGSNSAISACTHAGTQAKKSDG 1890  
Db 2031 TTTSDPGOSSITOGIPONDIKSL-NQVTTTTSSVQGVPSPPVVKVTKETSVDKOG 2089  
Qy 1891 TAGTTTTAGATCTVKGFF--AGOTAVGAVSVGASGAERRIQNVAAG-----EVSAT 1938  
Db 2090 TTRSSSTTTTTTTKGSNOGSLTLPLAVD--GLKSTKTTTTTKGTLKSLDLSLPEVDAS 2148  
Qy 1939 STDVANGSOLYKATOSIANATNELDHRHQNEN--KANAGISSAMAMASMPQAYIPGRM 1996  
Db 2149 I--AVNGDESRSASIKDTNLSKIDLSLPKLDASLVNNGKSSKSSSTTTTTTTSTKGNKV 2206  
Qy 1997 -----VTGGIATHNGQAVAVGLSK---LSDNGQWVFKINGSADTQ 2035  
Db 2207 SLSLPEVDASTAV--NGDDARSASIKDTNLSKIDLSLPKLDASLVNNG 2253  
RESULT 14  
Q55582 PRELIMINARY; PRT: 3029 AA.  
ID Q55582

Thu Sep 13 14:19:01 2001

AC Q5582; 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL 308.8 KDA PROTEIN.  
GN SLR0364.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PCC6803;  
RA Tabata S.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96127529; PubMed=8590279;  
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
Sugliura M., Tanaka S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
region from map positions 64% to 92% of the genome.";  
RL DNA Res. 2:153-166(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
Miyajima N., Hirose M., Sugliura M., Sasamoto S., Kimura T.,  
Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
Synechocystis sp. strain PCC6803. II. Sequence determination of the  
entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
DR EMBL; D63999; BAAL0087.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 3029 AA; 308789 MW; BEE0BB57E822A4FB CRC64;  
  
Query Match 5.08; Score 518; DB 2; Length 3029;  
Best Local Similarity 21.88; Pred. No. 2e-13;  
Matches 521; Conservative 256; Mismatches 901; Indels 710; Gaps 119;  
  
QY 56 VTGATL-----SGSAYA-----OKKDKTHIAIGE-----QKQPR 84  
Db 627 VSGATISNVSNGSVYTVTVTGIANNGTVNLDIFASGQNIIRDNTANNALSNTPTTDEQYT 686  
QY 85 RSGTAKAGD---DRAIAGENAN-AQGQATAI-----GSSN--KTVNGSSLDKIGTDAI 133  
Db 687 LONTAPASITILDANITADDDIINIAESGOAIPITGVGGEFNVGDTVTLTVNDKFTTGA 746  
QY 134 QGESI---AIGDDVKASDASIAIGSDDLHLLDGHGPKHPKGLINDLINGHAULKEI 189  
Db 747 GAGGLESINVPSSDLIVDADLIIA---SIATTDAGN-----L 782  
QY 190 RSSKONDVKYRRTA-----SGHASTAVGAMS----- 216  
Db 783 GSATDNGTVTDVTDAPITPIITVNDVTDADNIIINAEISGOAIPITGVGGEFNVGDTVTLTV 842  
QY 217 -----YAGHES-----NAFCTRATAKSAYSLAVGLAATAEGOSTIAIGSDA 258  
Db 843 NGPFTGTVDANGDFSIDVLGDLVNGSLDITAAVAITDAAAGNPGSASDNQTYTVDTTA 902  
QY 259 TSSSLGAIALG-----ACTRAQLQG-SIALGQGSVVVQOS--DNNSRP-----AYTNT 303  
Db 903 PTVTINAVDDDIINAVEAGSPVAVSGTTTGVGDDGVTVTIDGNTYTAIVTGNATWNI 962  
QY 304 QALD-PKFOATNTKACPLSIGNSIKRKLIINVGACVKNKTDAVNAQLEAVVVKWAKERI 362  
Db 963 PVADIANFEATEEVAVPVSVDLAGNPATRNII-----IVDIVAPA-----V 1004  
QY 363 TFGDDNSTDVK-----IGLDNTLTIKGAETNALTDNNIIGVVKWEADNSGLKVLAKTLNN 418  
  
Db 1005 TIDSISDDTGAQANDFTINDDTLVFNGTAEADST-----VVSVDG----- 1045  
QY 419 LFEVNTTTLNATTVTKVSSSSST---TAEILLSD-----SLTFTQPTNGSOSTSKTV--- 466  
Db 1046 -IEIGTVTAN-----GAGEWLIDYGTGLLAGDGYELSVTATNPTGNSATATQIVVDVT 1097  
QY 467 ----YGVNGVKFTNNAETTAIGTRITRIKIGFARDGDVDEKQAPYLDKKKQLKQKSVVAI 522  
Db 1098 TAPVTVINAVDDIINAVEAGSPVAVSGTTTG-VEGQV-----VTG 1139  
QY 523 TID-----NGIDAGNKKISNLAKSSANDAVTIEOLK-----AAKPTLNAGAGISVPT-- 571  
Db 1140 TIDGNTYATVTGNATFNIPVADIANFEATEEVAVTVDLAGNPATRNITVDTTAP 1199  
QY 572 -----EISVD-----AKSGNVT-----APTNYIGVKTTIELNSDGTGSKFVSK 608  
Db 1200 TVTINAVDDIINAVEAGSPVAVSGTTTGVGDDGVTVTIDGNTYATVTGNATWNI 1259  
QY 609 GSGTNNSLVTAHSLASYLNEVNRRTADSLAQSFVTKVEDDDANATV-----AKD 658  
Db 1260 VADIANFEATEEVAVTVDLAGNPATRNITV-----DTVAPAVTIDISDDTGAQAND 1315  
QY 659 TTKNAGAVSILKLGKGLTVATKDGTVTFGLSQDSGLTIGKSLNNDLGLTVKDTNEQI 718  
Db 1316 FITNDDTLVF-----NGTAEA---DSIVVSVL---DGEIGTVTANGAGEWLIDYTGL 1363  
QY 719 QVGANGIKFTVNGSNPGTGIANTRITRDKIGFAGSDCAVDTKPKYLDQKLOQVGNVKI 778  
Db 1364 L--ADGDYELSVTATNPG--NSATATQIV-----VDTTAPTVTINAVDDI-- 1408  
QY 779 TWYGINA--GGKAITGLSPTLPSIADQSSRIEL--GNTIQDKDKSNAASIN----DILNT 831  
Db 1409 ----INAVEAGSPV--AVSGTTTGVGDDGVTVTIDGNTYATVTGNATWNIPIVADIAN- 1462  
QY 832 GFNLKNNNPIDFVSTYDIV-----DFANGNATAT--VTHDTANKTSKVVDVWVDD--- 882  
Db 1463 -----FEATEEVAVTVDLAGNPATRNITVDTTAPT-VTINAVDDIIN 1509  
QY 883 -----TTIHLTGTDONKLGKVTTKLNTKTSANGTAT-----NFVNSSD----- 922  
Db 1510 AVEAGSPVAVSGT---TTGVGDDGVTVTIDGNTYATVTGNATWNIPIVADIANFEAT 1565  
QY 923 EDALVNAKDIAENLNTLAKEIHTTKGTADTALQTFVTKVKVDENNADNANATVCGKNNAN 982  
Db 1566 EEVAVTVDLAGNPATPA---TRNITVTVAPAVTIDISDDTGAQANDFTI----- 1614  
QY 983 NQVNTLTLKGELNKTIDKNGTVTFGINTTSLKAGKSLNDGGLSIKNPTGSEIQIOVG 1042  
Db 1615 -NDDTLVFNQ-----TAEADSTVVVSVLD---GIEIGTVTANGAGEWLIDYTGT---LL 1660  
QY 1043 ADGVKFAKVNNGVWAGAGIDGTRITRDEIGFTGTNGSLDKSKPHLSKDGINA--CGKKI 1100  
Db 1661 ADGDYELSVTATNPTGNSATATQIVVDTTAPTVTINAI-----AVDDIINAVEAGSPV 1714  
QY 1101 -----TNIOSEIAQ-----NSHDVATGKIIYDLKTELEN--KISSTAKTAQNSLHES 1147  
Db 1715 AVSGTTTGVGDDGVTVTIDGNTYATVTGNATWNIPIVADIANFEATEEVA-----T 1768  
QY 1148 VADSOGN-----NETVSN--PYSSYDTSKTSQVITF---AGE-----NGITTKVNGK-VV 1191  
Db 1769 VSDLAGNPATRNITVDTTAPTVTINAVDDIINAVEAGSPVAVSGTTTGVGDDGVV 1828  
QY 1192 RVGIDOTKGLTTPKLTGVGNNGKGI--VIDSQNGNTITGLSNLTANVTNDKGSVRTTEOG 1250  
Db 1829 TVTID---GNTYATVTGNATWNIPIVADIANFEAT-----EEVAVTVDLAGNPATRN 1881  
QY 1251 NIIDKEDKTRASTVDVLVSAGFNLOGGEAVDFVSTYDVTNFDAGNA---TTAKVTYDDT 1307  
Db 1882 NI--TVDIVAPAVTIDISDDTGAQAN---DFITNDDTLVF--NGTAEADSTVVVSLDGI 1934  
QY 1308 ---SKTSKVYDVVNDVDTTIEVKDKKLGKVTTLTTLTGTGANKFALSNOATGDALVKASD 1364



QY 1570 FVDSGOKANTPVL SANGLDLGGKVISNKGKTKD TDANVOOLNEVRNLLGLGNAGN- 1628  
 Db 1566 DDDSG--GDDTp---SDGGDGGNV-----TPPDDG-----GDGNGV 1698  
 QY 1629 ----DNADGNOVNIAIDIKDPNSGSSSNRTVIRKAGTVLGGKGNNDTEKLATGGI QVGVDK 1684  
 Db 1699 TPPDDGGDGGDTPPD-----HGGDVAPOYRADIGAYMGQWARNLQM-----QTLYDR 1748  
 QY 1685 DG----NANGDLSNVVVKTKDGSKKALLATYNAAGOTNY-----LTNNPAPAD 1730  
 Db 1749 EGSOYRNADG---SWARF-KAGKAESAVSGNIDMSNY SQFOLGGDILAWNGOOSV- 1803  
 QY 1731 RINBOGIRFFHVNDGNOEPVVOGRNGIDSSASGK-HSVAIGFQAKADGEAAVAIGBQTOA 1789  
 Db 1804 --TVGVMA SYINADTDSTGNRGADGSGQFTSSGNVDGYNLGVYATWFPADA-----QTHS 1854  
 QY 1790 G-----NOSTAIGDNAQATGQOSTAIGTGNVWAGKHSAGTCDPSTVKADNSYS 1837  
 Db 1855 GAYVDSWYQYGFYNNVESGDSGESYD-----STANNV-----SLETGYRYD 1897  
 QY 1838 VCNNNQFTDATQTDVFGVGNNTVTESNSVALGNSAISAGTHACTQAKKSDGTAGTT 1895  
 Db 1898 IALSN-----GNTVSLTPQAQVVMQNSADSVKADNYGTRIDGODGDSWTT 1942

Search completed: September 13, 2001, 12:55:10  
 Job time: 760 sec

QY 646 DDDANAITVAKDTTKN--AGAVSILKLGKNGLTVA---TKDGTVTFG--LSODSOLT 698  
 Db 725 DIDNSGLVAGGELENTLSGSGSLVK-TGTGELTSLGDNITYSGGTTITGGTTLTADHADS 783  
 QY 699 IGKSTLNNLGL-----TVKDTNEIOVANGIKRFTNVNGSNPGTGTANTARITRD 748  
 Db 784 LGSGLDINSGLVKVGEDELENTLSGSGSLVKTGTGELTSLGNDYSGGTTIIG-OTLAD 842  
 QY 749 KICFAGSGADVDPNKPYLDQDKLVGNKVTNTGKNAGKAITGLSP-TLPSIADQSSRN 807  
 Db 843 HADSLGT-GAV-----ANSGLVQGELENTLSGSGSLVKTGTGELTSLGDNYSYSGT 895  
 QY 808 IELGNTIQDKKNAASINDILNTGNLKNNNPIDFVSYDIDVFANGNATTAIVTHDT 867  
 Db 896 TIIIGGTL-----TADHADSLGTG-----AVANSGLVQGELENTLSGSGS 937  
 QY 868 ANKTSKVYDVNVDDTTIHLTGDDNKKLGKVTTKLNTKSANGNTATNFVNSDEDALV 927  
 Db 938 LVKTG-----TGELTSLGDNYSYSG-----GTTIIGGTLTADHADSLGTGAVA 979  
 QY 928 NAKDI-----AENLNTLAKEIHTTK-GTADTAL---OTFTVKKVDENNNADANAITVGOK 979  
 Db 980 NSGLVQGELENTLSGSGSLVKTGTGELTSLGDNYSYSGATTITDGTLLAANVNALGSG 1039  
 QY 980 NANNQVNTLILKGENG---LNIKTDKNGTVTFGINTSGLKAGKSTLNDGG-LSIKNPT 1034  
 Db 1040 NIDNS-CTLILDA--NAGFELANVTTHSGATTALAAST--LDAGQFTQEDGSLTSLDLGA 1095  
 QY 1035 GSEIOVAGADVGFPAKVNNGVVGAGIDGTRITRDEIGFT--GTNGSLDKSPHLSKDG 1092  
 Db 1096 ATDDAVITADSVALLGGTLN-VTGIG-SVTDSTWTPAYTYTLIGSDSAITTFDDDLTVAG 1152  
 QY 1093 INAGKKITNTQSGEI--AQNSHDVAVTGKLYDLKTELENKISTAKTAQNSLHEFSVAD 1150  
 Db 1153 MNREDDVDFLT-DGKVDADNTH-----YDL-----TASLWYADRDNATTD 1193  
 QY 1151 EOGNNFTVSNPYSSYDTSKT-SDV-----ITFAGENGIT--TK 1185  
 Db 1194 AHG-TFTLSDPGSGFNVAATLTDVDDTLDPGSRWDGKSLTKEGAGTLILSGDNDSYSGTT 1252  
 QY 1186 VNKGVVRY-----GIDOTKGLTTPKLTGVNNNGK- 1215  
 Db 1253 INEGLTVAASTALGTGLVDNNATLVLDADGEVSAVGITTHSGATT-QLALGTSLDLGD 1311  
 QY 1216 -----IVTDSONGNTITGLSNTLAN---VTNDKGSVRTTE-QGNI IKEDDKT 1259  
 Db 1312 SALIOQDGS TLNVELNSDSVQPLITGSSATLGGDLVSDASLOARASDAEFQSFKLMDMT 1371  
 QY 1260 RAASITVDVLSAGFNLOQNGEAVDFSVTYDVTNFDAG-----NATKAKTYDDTS 1308  
 Db 1372 SDIS-GDFTSLTMNLT---DKPDYLTVTCTINPEDASEYLLTEGLSWNATATSAT----- 1422  
 QY 1309 KTSKVYDVNVDDTTIEVKDKKLGKVT-----TTLTSTGTGANKFALSNOATGDALVK 1361  
 Db 1423 -PAHGTFTLIGAGD-SFEV-TSVLGDKTGNGDWDGKSLTKLGA KLTLSGANTYSYSDTNVQ 1479  
 QY 1362 ASDIHAHLMTLSGDIQTAK-CASQANNSAGYVDADGNKVIYDSTDNKYVQAKNDGTVDKT 1420  
 Db 1480 EGTL-----WLSGDSGISGMBGQQAVNVASGATFG-----SNGTTVNG 1518  
 QY 1421 KEVADKLVAAQATPDGTLAOMNKS-----VINKEOVNDANKKOGINE---DNAFVKG 1471  
 Db 1519 K-----VTNEGTLVFNSEETGAFTFNGDLINMGMTSGSSSTPGNTLYVDG 1567  
 QY 1472 LEKAASDNKTKNAA-----VTVGDNLNAQAQTLPTFAGD-TGTAKKLGELTIIK----GQ 1522  
 Db 1568 -----NYTGNGGSLYLTVLGDDDS-ATDKLVITGDASGTT-----DLYINGIGDGA 1613  
 QY 1523 TDTNKLTDNNIGVWACTDGTFLAKDLTNLNSVAG-----GFKIDDKGVS 1569  
 Db 1614 QTTNGIEVVDVGVSSTDAFELK-----NEVNAGLYTYRLWNSDNDWYLASKAQS 1665







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 22:50:01 ; Search time 33595.7 Seconds  
(without alignments)  
2835.657 Million cell updates/sec

Title: US-09-361-619-8  
Perfect score: 6159  
Sequence: 1 atgaatcacatataaagt.....ttggtgcagggttttcacttt 6159

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*

1:	gb_bal:*
2:	gb_ba2:*
3:	gb_ba3:*
4:	gb_in1:*
5:	gb_in2:*
6:	gb_in3:*
7:	gb_om:*
8:	gb_ov:*
9:	gb_pat1:*
10:	gb_pat2:*
11:	gb_ph:*
12:	gb_p11:*
13:	gb_p12:*
14:	gb_p13:*
15:	gb_p14:*
16:	em_bal:*
17:	em_ba2:*
18:	em_fun:*
19:	em_htgo_hum:*
20:	em_htgo_inv:*
21:	em_htgo_rod:*
22:	em_htg_hum1:*
23:	em_htg_hum2:*
24:	em_htg_hum3:*
25:	em_htg_hum4:*
26:	em_htg_hum5:*
27:	em_htg_hum6:*
28:	em_htg_hum7:*
29:	em_htg_hum8:*
30:	em_htg_inv1:*
31:	em_htg_inv2:*
32:	em_htg_other:*
33:	em_htg_rod:*
34:	em_hum1:*
35:	em_hum2:*
36:	em_hum3:*
37:	em_hum4:*
38:	em_hum5:*
39:	em_hum6:*
40:	em_hum7:*
41:	em_in:*
42:	em_om:*
43:	em_or:*

44: em\_ov:\*

45: em\_pat:\*

46: em\_ph:\*

47: em\_pl:\*

48: em\_ro:\*

49: em\_sts:\*

50: em\_sy:\*

51: em\_un:\*

52: em\_vi:\*

53: gb\_sts1:\*

54: gb\_sts2:\*

55: gb\_sts3:\*

56: gb\_sy:\*

57: gb\_un:\*

58: gb\_vil:\*

59: gb\_v12:\*

60: gb\_htg1:\*

61: gb\_htg2:\*

62: gb\_htg3:\*

63: gb\_htg4:\*

64: gb\_htg5:\*

65: gb\_htg6:\*

66: gb\_htg7:\*

67: gb\_htg8:\*

68: gb\_htg9:\*

69: gb\_htg10:\*

70: gb\_htg11:\*

71: gb\_htg12:\*

72: gb\_htg13:\*

73: gb\_htg14:\*

74: gb\_htg15:\*

75: gb\_htg16:\*

76: gb\_htg17:\*

77: gb\_htg18:\*

78: gb\_htg19:\*

79: gb\_htg20:\*

80: gb\_htg21:\*

81: gb\_htg22:\*

82: gb\_htg23:\*

83: gb\_htg24:\*

84: gb\_htg25:\*

85: gb\_pr1:\*

86: gb\_pr2:\*

87: gb\_pr3:\*

88: gb\_pr4:\*

89: gb\_pr5:\*

90: gb\_pr6:\*

91: gb\_pr7:\*

92: gb\_pr8:\*

93: gb\_pr9:\*

94: gb\_rol:\*

95: gb\_rod:\*

96: gb\_in4:\*

97: gb\_pr10:\*

98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	6159	100.0	6159	10	AX079920 Sequence
2	4101.6	66.6	6141	10	AX079918 Sequence
3	4101.6	66.6	6972	10	AX079917 Sequence
4	4090.6	66.4	6973	9	AR040716 Sequence
5	4090.6	66.4	6973	10	AX079913 Sequence
6	3982	64.7	62909	10	AX067457 Sequence
7	3947.8	64.1	5976	10	AX079914 Sequence
8	3946.8	64.1	6259	10	AX079924 Sequence

Thu Sep 13 14:19:03 2001

AX079957 Sequence  
AX079922 Sequence  
AX079959 Sequence  
AC007926 Trypanoso  
AJ277635 Pasteurel  
AE006194 Pasteurel  
I55122 Sequence 1  
U38617 Haemophilus  
U38657 Escherich  
U41852 Haemophilus  
U32846 Haemophilus  
AJ277623 Salmonell  
AC084397 Trypanoso  
U67264 Helicoverpa  
AF271059 Helicover  
Z98549 Plasmodium  
Z98547 Plasmodium  
AE003163 Drosophil  
AC007926 Trypanoso  
AE003782 Drosophil  
AE003982 Xylella f  
AC006279 Plasmodiu  
AC005140 Plasmodiu  
AL049184 Plasmodiu  
AC013254 Drosophil  
X02426 Trypanosoma  
AC00125 Drosophil  
AF051898 Dictyoste  
I55123 Sequence 3  
AL049185 Plasmodiu  
Continuation (2 of  
AF201315 Dugesia p  
AF18151 Dictyoste  
AJ277636 Pasteurel  
AE006108 Pasteurel  
AC006281 Plasmodiu

ALIGNMENTS

RESULT 1  
LOCUS AX079920 6159 bp DNA PAT 22-FEB-2001  
DEFINITION Sequence 8 from Patent WO0107619.  
ACCESSION AX079920  
VERSION AX079920.1 GI:13159440  
KEYWORDS  
SOURCE Moraxella catarrhalis.  
ORGANISM Moraxella catarrhalis  
Bacteria: Proteobacteria; gamma subdivision; Moraxellaceae;  
Moraxella.  
1 (bases 1 to 6159)  
Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.  
Recombinant high molecular weight major outer membrane protein of  
moraxella  
Patent: WO 0107619-A 8 01-FEB-2001;  
CONNAUGHT LABORATORIES LIMITED (CA)  
Location/Qualifiers  
i..6159  
/organism="Moraxella catarrhalis"  
/db\_xref="taxon:480"  
BASE COUNT 2035 a 1386 c 1385 g 1353 t  
ORIGIN  
Query Match 100.0%; Score 6159; DB 10; Length 6159;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 atgaatcacatctataaagtcatctttaacaagccacagccacatttatggcgggagc 60  
|||||

1 ATGAATCACATCATATAAAGTCATCTTTTAAACAAAGCCACAGGCACATTTATGGCCGTGGCG 60  
61 gaatatgccaaatcccaacagtagcaggggggggagtagctgtctacaggggcaagtggcagt 120  
|||||  
61 GAATATGCCAAATCCACAGTAGTACGGGGGGGGGTAGCTGTGTCTACAGGCAAGTTGGCAGT 120  
|||||  
121 gtacgcacttaagcttttcccgctattgcgcgcgtctgcgtctctctgcgtatcgggtgcagc 180  
|||||  
121 GTACGCACACTTAAGCTTTGCCCGCTATTGCCCGCGCTCGGTCTCTCTGCATCGGTGCGAGC 180  
|||||  
181 ctcaatggcagtgcttctctcaacaaatctactacaagatcgaaatgggtcaacaacaac 240  
|||||  
181 CTCAATGGCAGTGCTTATGCTCAACAAATACTTACCAAGATCGAAATGGTCAACAAAC 240  
|||||  
241 aacataaacaacacgctgaagcgatgccttagcgacagcaggtgaagcatccattgctttt 300  
|||||  
241 AAGATAAACAAACGCTGAAGCGGATGCCCTAGCAGCAGGTGAAGCATCCATTCGCTTT 300  
|||||  
301 ggtagctcttcttaagcacaagcgtctcaagctattgctatcgttagtgcgttagtcaaacagat 360  
|||||  
301 GGTAGTCTTTCTAAGGCACAAAGGCTCTCAAGCTATTGCTATCGGTAGTGTCAAAACAGAT 360  
|||||  
361 cctaataatggtagtgaatgaatgtaggttccacgcgcaaaaggttaacagtagtcacgcgc 420  
|||||  
361 CCTAATAATGGTAGTAATGTAATGTAAGTTCCACGCCCAAGGTAAACGAGTCCATTCGCT 420  
|||||  
421 atcgggtgtagtattgggtgaggggtgtagtcgtcgatgctcgttagtgcgttagtgcgttca 480  
|||||  
421 ATCGGTGTTGATGATTGTTGGCTGAGGGTGATGCCCTCGATTCGCTATCGGTAGTGTCAAAAC 480  
|||||  
481 tatttgcttaagaatctgtatctgaagaatgaatttcacaaacttatctatgcccattgaa 540  
|||||  
481 TATTGTTGCTTAAGATCTTGTATCTCAAGAAATGAATTCACAAACTTATTATGTCATG 540  
|||||  
541 atattaaaaaaatacacaacccacccgtaggtgtaaaatcaaaatcaaaatcaaaatcaaaagca 600  
|||||  
541 ATATTAAAAAAAATAACAAACCTCAACCGATGTAATAATCAAAATATCAGCGCAGCAAGACCA 600  
|||||  
601 caagggcagccagctagctagctgagggagccatgcatatgcacaggggtcatttttccaac 660  
|||||  
601 CAAGGGCAGCCAGTACTGACAGTGGCAGCCATGTCATATGCACAGGGGTCAATTTTTCGAAC 660  
|||||  
661 gcttttggacatacagcaacagctgaagctgcttcttctgagcagtaggttcgctgagcgc 720  
|||||  
661 GCCTTTGGTACATACGCAACAGCTGAAGCTGCCTATCTCTGGCAGTAGGTCTTGGCCGCC 720  
|||||  
721 caagccacaacaaatcttcaatcgctgtggttcccaatgcaaaagctaaacgcttgcga 780  
|||||  
721 CAAGCCACAAACAAATCTTCATTCGCTGTTGTTCCCAATGCAAAAGCTTAAGCGGCTTGCA 780  
|||||  
781 ggcagcagccattggttggaataactgtagttaatttgggtcgagggcgttcgcttaggtttt 840  
|||||  
781 GGCACAGCCATTGGTGAATAACTCTAGTTAATTTGGGTGAGGGGCTTGCCTAGCTTTT 840  
|||||  
841 ggttcagatccttgaatgggataataacagatgagcagtgccctatgaccactaggt 900  
|||||  
841 GGTTCAGATCCTTATGATGGGATAATAATACAGATGCCAGTGCCTATGTACCACTAGT 900  
|||||  
901 aaaacgttagcagaccagtagtaaaagccaccccgaggggtgattctacagcatatatttcc 960  
|||||  
901 AAAACGTTAGCAGACCCAGTATAAAGCCACCCGCCAGGGGTGATCTACGGATATATTTTCC 960  
|||||  
961 attggttaataataataataacagtagcagctcagcgttaaaatcacaatcgcgtgagcgc 1020  
|||||  
961 ATTGGTAATAGTAATAATAATAGCAGTATCAGCGCGTAAATCATCAATGTGCGTGGC 1020  
|||||  
1021 ggttcgagatcccgatcggttcaatggtgcacagcttaaatggttgaggaactggct 1080  
|||||  
1021 GGTTCGAGATCCGATCGGTCGAATGTGCGCAGCTTAAATTTGGTGGAGGAACCTGGCT 1080  
|||||  
1081 aatcgtaaaatctatttaaggggtgagtgagcaataatagcgaatagcgttagaagaaggt 1140  
|||||  
1081 AATCGTAAATTTACTTTTAAGGGTGATGGTGACACANTTAATAGCAATAGCGTAGAAGAGGT 1140  
|||||

Qy 1141 ttgggcaatacctttaactatttaaaggtagtcacagaccacacgattaaacgaagctaac 1200  
Db 1141 TTGGGCAATACTTAACTATTAAAGGTTGATGCACAGACCACACGCGATTAAACGAAGCTAAC 1200  
Qy 1201 atcgggtggtaacagatggcaatgggtcctgaaagttaaaactgtctaaagaactgaactgga 1260  
Db 1201 ATCGGTTGGTAAACAGATGGCAATGGTCTGAAAGTTAAACTTGTCTAAAGAGCTACTGGA 1260  
Qy 1261 ttgaccagtgctccgctaccacaaaatccaccgtttagtaataaccaacaacaacgaagcc 1320  
Db 1261 TTGACCAGTGCTCCGCTACCAACAAAATCACCGTTAGTAAATACCAACAACAACGCGC 1320  
Qy 1321 gagctaaaaagcgggtggttgacotttagcccaataacaggtacacaaacagataaaaccc 1380  
Db 1321 GAGCTACAAAGCGGTGGTGTGACCTTAAAGCCAAATACAGCGTACAAAACAGATAAAACC 1380  
Qy 1381 gtctacagcattgatgattgaatttactaataatagtaataatagtaataatagtaataaaggt 1440  
Db 1381 GTCTACAGCATTGATGATGAACTTACTAATGATAGTATAGTAAGTAAGTAAGGT 1440  
Qy 1441 actactgattaccacaaaagaattggtttgtgctgactaataatgagggagttgatgaa 1500  
Db 1441 ACTACTGATTACCAAAAAGAAATGGTGTGCTGCTACTAATGATGAGTGTGATGAA 1500  
Qy 1501 agcaaaccttcttgacaacgaagaactaaaagtgtggcaacagaccccttaaacagtggt 1560  
Db 1501 AGCAAAACCTTATCTTGACAACGAAGCTAAAGTTGGCAACAGCACCCCTTAACAGTGGT 1560  
Qy 1561 agcttgactgttaatacacaccactggtaataaacaacatccaagtcggtgcttaagtgatt 1620  
Db 1561 AGCTTGACTGTGTTAATACACCACCTGGTAATAAACAATCCAAAGTCGGTGTCTAATGGCAATT 1620  
Qy 1621 aaattggccacagtcgctaaatggttgcaaataccctcagcaacagtcggcactgctcgt 1680  
Db 1621 AAATTGGCCACAGTCGCTAATAATGTTGCAATACCTCAGCAACAGTCGGCACCTGCTGTT 1680  
Qy 1681 attaccgaagagaaaattggttggctgactaataatgatgagttgatagaacaagcacc 1740  
Db 1681 ATTACCGGAAGAGAAAATTGGTTTGGTACTAATGATGGAGTTGATGAACAAGCACCA 1740  
Qy 1741 tatctgataaagaacactttaagtggtgctgtttgaaattacacagatagtggtatt 1800  
Db 1741 TATTGGTAAGAAGACGACTTAAAGTGGTGTGTTGAAATTTACACAGATAGTGGTATT 1800  
Qy 1801 aatgctgtaatacaagaattaccggacttactaatggtatagcaaaatcccgatgcggtt 1860  
Db 1801 AATGCTGGTAATCACAGATTACCGGACTTACTAATGCTATAGCAAAATACCGATCGCGTT 1860  
Qy 1861 accatcaaacagctcaaaagacgcaagcctactttaaacgagcgatggcatcagatt 1920  
Db 1861 ACCATCAAAACAGCTCAAGACGCCAAGCCTACTTTAAACGCGAGCGATGGCATCAGTATT 1920  
Qy 1921 aatgataaacggggtcagttgattagtagtggaataattaccaccccaacttaaac 1980  
Db 1921 AATAGTAATAACGGGGATCTAGTTAGTAGTAGTGGCAATATTACACCCCAACTTATAAC 1980  
Qy 1981 attagctgaaacacactaaagcttaacagtaattggcaccagtggttaataataaatttagt 2040  
Db 1981 ATTAGCGTGAACACCACTAAGCTTAACAGTAATGGCACCGTGTGTAATAATAAATTAGT 2040  
Qy 2041 gttagtaatgctatgataaacaatagcttagttaccgccaagaatttggcagactatcta 2100  
Db 2041 GTTAGTAATGCTCATGATGAACAATAGCTTAGTTACCGCCAAAAGATTTTGGCAGACTATCTA 2100  
Qy 2101 aataaagtcaatgaaacggctgacagtgctctaccaagctttaagtcacaaacgggtat 2160  
Db 2101 AATAAGTCAATGAACAGGCTGACAGTGTCTCTACCAAGCTTTAAAGTCCAAACGGGTAT 2160  
Qy 2161 aatagcaaacacgcacatcacggtgggttaagaatacaaacggcagaccttcaacacctta 2220  
Db 2161 AATAGCAACAACGCCATCACCGTGGTAAAGATACNAAGCGCAAGACCTTCAACACCTTA 2220

Qy 2221 aaactcaaaagtgaaacacgggtgtttaattacgaccaatagagccacagggtacagttacc 2280  
Db 2221 AAACTCAAAGGTGAAACCGGTGTTAATATTACGACCAATAGAGCCACAGGTACAGTTACC 2280  
Qy 2281 ttggcattgaccaaaagtaatgggtctccacgctcaccgcttaagctgacggtgggtagcagataca 2340  
Db 2281 TTTGGCATTGACCAAAAGTAATGGTCTCACACGCCCTAAGCTGACCGTGGGTAGCGGATACA 2340  
Qy 2341 aatgtaatacgtattggtgttattgagcaagtcctcctagcgtgacggttaacagcaccacaaaac 2400  
Db 2341 AATGTAATCGATTGGTTATTGAGCAATCCCTAGCGCTGACGGTAAACAGCACCAAAAAC 2400  
Qy 2401 atcattaaagattgtcccccacactgctcagcattgcccagtcctcaagtggcgcaacata 2460  
Db 2401 ATCATTAAGGATTGTCCCCCAACACTGCTTAGCATTTGCCAGTCCCAAGTGGCGCAACATA 2460  
Qy 2461 gcactgggcatacaatcgaagaaaaagacaaatccacgctgcccagctgacgattgatgatg 2520  
Db 2461 GCATGGGCAATACNATCGAAGAAAGACAAATCCACGCTGCCAGCATTTGATGATGTG 2520  
Qy 2521 ctaaatgcaggcttttaacctaaaaataatggcaagacaaagactttgtctccacttat 2580  
Db 2521 CTAAATGCAGGCTTTTAACCTAAAAAATAATGGCAAGACAAAGACTTTGTCTCCACTTAT 2580  
Qy 2581 gacactgtgactttatcgatggcaatgccaccacccacagtaacttatgatgaagcc 2640  
Db 2581 GACACTGTTGACTTTATCGATGGCAATGCCACACCGCACACAGTAACTTATGATGAAGCC 2640  
Qy 2641 aatcaaacccagtaagtggtgctatgatgaatggtgagatgagaacacattgaactgaca 2700  
Db 2641 AATCAAACCCAGTAAGTGGCGTATGATGTAATGTGGATGAGAAACCAATTTGAACGTACA 2700  
Qy 2701 ggcgataatggcagaagaacacttgctcaaaacccatcaaaactgaccgaaacagtgact 2760  
Db 2701 GCGGATAATGCGAAGAAACAACCTTGGCGTCAAAACCATCAAACTGACCGAAACAAGTACT 2760  
Qy 2761 aatggtaagcaactacattagtagccagcagatgaccatgccccttggttaaagccagtgat 2820  
Db 2761 AATGGTAATGCAACTACATTTAGTACCAGCATGACCATGCGCTTTGTTAAAGCCAGTGAT 2820  
Qy 2821 atcgcggcaatcacaacaccctagcgagaaattcacaccacacacacacacacacacac 2880  
Db 2821 ATCGCGGCAATCTTAACACCCCTAGCCGAGGAATTCACACCACCAAGGCGACAGCAAC 2880  
Qy 2881 accgacctacaaacacctttaccgttaaaaggtagatgataaaatgataaggtgtagtacacc 2940  
Db 2881 ACCGCGCTACAAACCTTTACCGTTAAAAAGGTAGATGATAAATGATAAGCTGATGACACC 2940  
Qy 2941 aacgccatcacccgtgggttaagatggcacaagtggtaaagtcacaacaccttaaaactcaaa 3000  
Db 2941 AACGCCATCACCGTGGGTAAAGATGGCACAAAGTGTAAAGTCAACACCTTTAAACATCAA 3000  
Qy 3001 ggtaaaaacggctcttgatattaaaaccgacaaaagtgtagcgtttacctttggcattaac 3060  
Db 3001 GGTAAAAACGGTCTTGATATTAAAAACCCGACAAAGATGGTACGGTTACCTTTGGCATTAAC 3060  
Qy 3061 acccaaacggctcttaaacggcgacagcacactctaaacacaaatggccttgcctatt 3120  
Db 3061 ACCCAACGGCTCTTAAAGCGCGGCACAGCACCTCTTAAACAACAATGGCTTGTCTATT 3120  
Qy 3121 aaaaaacacgcgtagtaacgaacaaatccagtcggtgctgtagtggcgtgaagtttgcattg 3180  
Db 3121 AAAAACACCGCTAGTAAACGAACAAATCCAAAGTCGGTGTGATGGGTGAAAGTTTGGCATG 3180  
Qy 3181 gtaataatggttggtagtgctgagcattgatggcacaactcgcattaccagagatgaa 3240  
Db 3181 GTTAATAATGGTGTGTAGGTGCTGGCATTTGATGGCAACCTCGCATTTACCAGAGATGAA 3240  
Qy 3241 attggcttactgggactaataatggctcacttgataaaagcaaaccccccacctaagcaagac 3300  
Db 3241 ATTGGCTTTACTGGGACTAATGGCTCACTTGATATAAGCAAAACCCACCTTAAGCAAGAC 3300  
Qy 3301 ggcattaacgcagggtgggtaaaaagattaccaacattcaatcagggtgagattgccccaaac 3360

Db	3301	GGCATTAAACGAGGTGGTAAAAAGATTACCAACATTCATCAGGTGAGATTGCCAAAAC	3360
Qy	3361	agccatgatcgtgtgacagcgccgaagatttatgtatttaaaacccgaacttgaaataaa	3420
Db	3361	AGCCATGATGCTGTGACAGCGGCAAGATTATGATTTTAAACCCGAACCTTGMAAATAA	3420
Qy	3421	atcagcagtlactgcgaacacagacacaaactcattacacgaattctcagtagcagatgaa	3480
Db	3421	ATCAGCAGTACTGCCAAAACACACAAAACCTCATACACGAATTCATCAGTAGCAGATGAA	3480
Qy	3481	caaggtaataacttttcggttagtaaaccttactccagttatgacacctcaagacctct	3540
Db	3481	CAAGGTAATAACTTTACGGTTAGTACCCCTTACTCCAGTTATGACACCTCAAGACCTCT	3540
Qy	3541	gatgtcctacacctttgcagtgaaacgcaattaccaccaaaggttaataaaagtggtg	3600
Db	3541	GATGTCATCACTTTGCAGGTGAACACGGCATTTACCACCAAGTAAATAAAGGTGTGGTG	3600
Qy	3601	cgltgtggccattgaccaaaccacaaagcttaccacagccctaaagctgacctggtaataat	3660
Db	3601	CGTGTGGGCATTGACCAAAACCAAGGCTTAACACACGCCCTAAGCTGACCGTGGTAAATAAT	3660
Qy	3661	aatgcgaagcattgttcattacagcgaataatggtcgaataatcaccatcacaggaactaagc	3720
Db	3661	AATGGCAAGGCATTGTTCATTACAGCCAAATGGTCMAAATACCATCACAGGACTTAAGC	3720
Qy	3721	aacactctagtaattgtacccatgataaaggtagcagccacacgaacgaagggcaat	3780
Db	3721	AACACTCTAGCTAATGTACCAATGATAAAGTAGGTAGCGTACGCCACACAGAGGCAAT	3780
Qy	3781	ataatcaaaagcagaacaaacccgtgcgcagcagattgtgatgtcctaagcgcaggc	3840
Db	3781	ATAATCAAAAGCAGAAACAAACCCGTGCCGCCAGCACTGTTGATGTGCTTAAGCGCAGGC	3840
Qy	3841	tttaacttgcgaagcaatggtaagcgggttgactttgtctccacttatgacacgcgtcaac	3900
Db	3841	TTTAACTTGCAAGCAATGGTGAAGCGGTTGACTTTGCTCTCCACTATATGACACCGTCAAC	3900
Qy	3901	tttgcgaatggcaatacacacacgcgttaagtgacctatgatgacacaaagcaaacagct	3960
Db	3901	TTTGGCAATGGCAATFACCAACACCGCTANGGTGACCTATGATGACACACAGCAAAACAGT	3960
Qy	3961	aaagtgttctatgataatgtgatatacaacacattgaagttaagataaaataaactt	4020
Db	3961	AAAGTGGTCTATGATGTCAATGTGATGATACAAACCACTTGAAGTTAAAGATAAAAACTT	4020
Qy	4021	ggcgtaaaaaacacacacattgaccagtagtcgacaggtgctaataatttgccctaaagc	4080
Db	4021	GGCGTAAAAAACCCACACATFACCACTGACTGGCACAGGTGCTATATTAATTTGCCCTAAGC	4080
Qy	4081	aatcaagctactggcgatgcgcttctgaagccagtgatctcgttgcctcatctataaacacc	4140
Db	4081	ATCAAGCTACTTGGCGTGGCTTGTCAAGGCCAGTGATATCGTGTCTCATCTATAACACC	4140
Qy	4141	tatatcggcgacatccaaactgcgaagggcaagccaaagcaacacactcagcaggtcat	4200
Db	4141	TTATCTGGGGACATCCAACTGCCAAAGGGCAAGCAAGCGAACTCACTACAGGCTAT	4200
Qy	4201	gtgatgctgatggcaataaagtcattctatgacagtlaccgataaacaagtaactataagcc	4260
Db	4201	GTGGATGCTGTATGGCANTTAGGTCATCTATGACAGTACCGATACCACTACTATCAAGCC	4260
Qy	4261	aaaaatgatggcacagtgtgataaaaccaaagaagttgccaagacaaactggtcgcccaa	4320
Db	4261	AAAAATGATGGCACAGTTGTATAAAACCAAGAGGTTGCCAAAAGACAACTTGGTCGCCCAA	4320
Qy	4321	gcccaaacccagatgacacattgctcaaatgaatgtaacatcagtcatttaacaaagaa	4380
Db	4321	GCCCAAAACCCAGATGACATTTGGTCAAAATGAATGTCAATCAGTCATTAACAAGAA	4380
Qy	4381	caagtaaatgatgccataaaagcaagggcatcaatgaagacacacgccttctttaaagga	4440
Db	4381	CAAGTAATATGATGCCAATAAAAAAGCAAGGCATCAATGAAGACAAACGCCCTTTCTTAAGGA	4440
Qy	4441	cttgaaaaagccctctctatacaaaaaacccacacccgcgcagtaactgtggtgatttca	4500
Db	4441	CTTGAAAAAGCCCTCTCTGTATAACAAAAACCAAAACGCCGCAGTAACCTGTGGGTGATTTA	4500
Qy	4501	aatgcgctgtcccaaacacccgctgacctttgcaggggatatacagggcacacccgctaaataa	4560
Db	4501	AATGCCGTTGCCCAACACCCGCTGACCTTTGCAGAGGGATACAGGCACAAACCGCTAAAAAA	4560
Qy	4561	ctggcgagacttttgaccatcaaaagtgaggcaaacagacacccaataagctaaaccgataat	4620
Db	4561	CTGGCGAGACTTTTGACCATCAAAAGGTGGCAAAACAGACACCAANTAGCTAACCGGATAAT	4620
Qy	4621	aacatcggtgtgtagcaggtactgtactgtcactgtcaaaacttgcgaagacccctaaacc	4680
Db	4621	AACATCGGTGTGGTAGCAGTACTGTACTGTCTCAAACTTGCCAAAGACCTTAACC	4680
Qy	4681	aatcttaacagcgttaatgacaggtggcaccacaaatgataaaaagcactctcttttga	4740
Db	4681	AATCTTAACAGCGTTAATGACAGGTGGCACCACAAATTTGTAAGTGCCAATGGCTGGACCTG	4740
Qy	4741	gacgcaaacggtcaagccaaagcaaacccctgtcgtctgaagtgcacatgggctggacctg	4800
Db	4741	GACGCAAAACGGTCAAGCCAAAGCAAAACACCCCTGTCTGTAAAGTGCCAATGGCTGGACCTG	4800
Qy	4801	ggtggcaaggttcacgttaatgtgtggcaagggcacaagaatataccgacgtgcacatgta	4860
Db	4801	GGTGCAAGGTCATCAGTAATGTGGCAAGGCACAAAGATACCGACGCTGCCAATGTGA	4860
Qy	4861	caacagttaaacgaagtagcgaactgtgtggcttggtaagtataacgcgtgacggcgaat	4920
Db	4861	CAACAGTTAAACGAAGTAGCAACTGTGTGGTCTTGGTAAATGATAGCGTACACGGCAAT	4920
Qy	4921	caggttaacattggcgcacatacaaaaagacccaaattcaggttcatctataacgcacact	4980
Db	4921	CAGGTAACATTGGCGACATCAAAAAGACCCAAATTCAGGTTTCATCATCAACCGCACT	4980
Qy	4981	gtcatcaaaagcagcagcgttactgtgcgttaaggtataacgataccgaaaaaactgtcc	5040
Db	4981	GTCAATCAAAAGCAGGCACGGTACTTGGCGGTAAAGGTAAATACGATACCGAAAAACTTGC	5040
Qy	5041	actgtgtgtgtacaaagtggtgcgtgataaagcggcgaacgcgtacacggcgttaagcaat	5100
Db	5041	ACTGTGTGTGTACAAAGTGGCGCTGGATAAAGACGGCAACCGCTAACGGCGATTTAAGCAAT	5100
Qy	5101	gttgggttcaaaacccaaagatggcagcaaaaagccctgcctgcctactataacgc	5160
Db	5101	GTTTGGGTCAAAACCCAAAAGATGGCACAANAANAAGCCCTGCTGCCACTTTATAACGCC	5160
Qy	5161	gaaggtcagacacaaactatgtgaccacaaaccccgagagggccattgacagaataatgaa	5220
Db	5161	GCAGGTACAGACCAACTATGTGACCAACAACCCCGCAGAAAGCCATTGACAGAATAAATGAA	5220
Qy	5221	caagggtatccgcttcttccatgacagatgcaatacaagagcctgtgtgtaaaagggct	5280
Db	5221	CAAGGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCCCTGTGTACAAGGCCGT	5280
Qy	5281	aacggcattgaactcaagtcgctcaggaagcactcagtggtgcgataagtttccagggcgaag	5340
Db	5281	AACGGCATTCGACTCAAGTGCTTCAGGCAAGCACTCAGTGCGGATAGGTTTCCAGGCCAAG	5340
Qy	5341	gcagatggtgaaagcccgcttgccataggcagacaaaccccaagcaggaacacacatccatc	5400
Db	5341	GCAGATGTTGAAGCCGCGTGTGCCATAGGCAGACAAACCCCAAGCAAGCAACCAANTCCATC	5400
Qy	5401	gcaatcggtataacgcacacgaacacggcgatcaatccatccatccatccatccatccat	5460
Db	5401	GCCATCGGTATACGCACAAAGCCACCGGCGATCAATCCATCCATCGCATCGGTACAGGCAAT	5460
Qy	5461	gtgtagcaggtgaagcactctgtgtccatcggaaccccaagcactgttaaggtcgataac	5520
Db	5461	GTGGTAGCAGGTGAAGCCTCTGTGTGCCATCGGCACCAAGCACTGTGTAAGGCTGTATAC	5520



Qy	1063	ttggtgaggaactggcta---atcgtaaaattacttttaagggtgatggtgacataaat	1119
Db	1054	CGCGTGTGTGAAGTGGGTAAAGCGGTAGAAATTTACTTTTCAGGGTGATG-----ATAAC	1107
Qy	1120	agcaatagcgtagaaagaggtttgggcaataacttttaactattaaaggtgatgcacagacc	1179
Db	1108	AGTACTGACGTAAANATAGGTTTGATATACTTTAACTATTAAAGGTGGTGCAGAGACC	1167
Qy	1180	aacgattaaaccgaagctaacatcgtgtgtgtaacagatggcaaa-----tggtctgaaa	1233
Db	1168	AACGCATTAAACCGATAATAATATACGGTGTGGTAAAGAGCGCTGATAATAGTGGTCTGAAA	1227
Qy	1234	gttaaacctgcataaagagctgactggattgaccagtgctccc-----gct	1278
Db	1228	GTTAAACCTTGCTAAACACTTTAAACAATCTTACTGAGCTGAATACAACATACATTTAAATGGC	1287
Qy	1279	accacaacaaatcacocgttagttaataccaacaacaaacgccgagctacaagcgtggt	1338
Db	1288	ACAACCACAGTTAAGGTAGTAGTAGTAGTACTACAGCTGAATTTATTGAGTGATAGT	1347
Qy	1339	ttgaccttagccca-----atacaggttcaaaaacagataaaaacgcgtctcacgact	1392
Db	1348	TTAAACCTTTTACCACGCCCAATACAGGCAGTCAAGGCACAGCAAAACCGTCTATGGCGTT	1407
Qy	1393	gatggattgaagtttactaatgatagtaatagtatagcaactaaaggtactactcgtatt	1452
Db	1408	AATGGGTCAAGTTTACTAAATACGAGAAACACAGCAGCAATCGGCACCTACTCGTATT	1467
Qy	1453	accaaaaagaaattggttttctggtactaatgatggagtgtatgaagcaaaccttat	1512
Db	1468	ACCAGAGATTAATTTGGCTTTGCTCGAGATGCTGATGTTGATGTAACCAAGCACCATAT	1527
Qy	1513	cttgacaacgaaagctaaaagttggcaacagcaccttaaacagtggttagcttgctt	1572
Db	1528	TTGGTATMAAACAACCTTAAAGTGGGTAGTGTGCAATTTACCATAGACAATGGCATGTAT	1587
Qy	1573	aaatacacccactggttaataaaacaaalccaagtcgctgcttaatggcattaaatlqccaca	1632
Db	1588	GCAGGTAAATAAAAAGATCATGTAATCTTGCCAAAGGTAGCAGTGTCTAACGATCGGTTACC	1647
Qy	1633	gtcgcttaaatgttgcataactcagcaacagtcggcactgctcgattaccogaagag	1692
Db	1648	ATCGACAGCTCAAAAGCCGCAACGCTACTTTAAACGCGAGCGCTGGCATAGTGTACACA	1707
Qy	1693	aaaattggttttgcgtactaatgatggagtgtatgaacaagcaccataatttgataaa	1752
Db	1708	CCTACTGAATATACAGTTGATGCTAGAGTGGCAATGTTTACCGCCCAACTTACAACA--	1765
Qy	1753	gaacgacttaaaagggtgctgttgaaattaccacagatagtggttattaatgctggtaat	1812
Db	1766	-----TTGGCGTGAAACACCCGAGCTTAAACAGTGTAGGCACCTAGTGATATAATTTAGT	1818
Qy	1813	cacaagattaccggaacttactaatatggatatagcaaataccgatgcggttaccattcaaacag	1872
Db	1819	GTTAAGGTAGTGTGTACGAACAATATGTTAGTTACCCCGCAACATTTGGCAAGCTATCTA	1878
Qy	1873	ctcaaaagacgcaaaccttaacttttaaacgcgagcgatggcaatcagttattaatgataaac	1932
Db	1879	AATGAAGTCAATCGNACGGCTGCAGTGTCTACAAAAGCTTTACCGTTTAAAGAAGAAGAC	1938
Qy	1933	ggggatctagttgatatagtagtggccaatataaccacccaaactataaacattagcgtgaaa	1992
Db	1939	GATGATGACG-----CCAACGCTATCACCGTGGCT	1968
Qy	1993	accactaagcttaacagtaattggcaccagtggttaataataaatttagttagtaatgct	2052
Db	1969	AAGATACGACAAAAAATGCCGGCGCAGTCAGCATCTTAAACCTCAAAAGGTAATAAACGGT	2028
Qy	2053	catgataacaaatagtttagttaccgcgaagaatttggcagactatcaataaagtcact	2112
Db	2029	CTAACGGTTGCTTACCAAAAAGATGGTACGGTTACCTTTGGGCTTAGCCAAATATACGGT	2088

Qy	2113	gaaacggctacagtgctctaccaaagcttttaaagtcacaaacgggtgataataatgacacaac	2172
Db	2089	CTGACCATTTGCAAAAGCACCCTAAACAACGATGGCTTGACTTTAAAGATACCAACGAA	2148
Qy	2173	gccatcacctgggtgaaagatacaaaacggaagacaccttaaaccttaaacctcaaaagt	2232
Db	2149	CAATCCAAAGTCGGTGAATGGCATTTAAATTTACTTAATGTGAATGGTAGTAAATCCAGGT	2208
Qy	2233	gaaaacgggtgtaattattagacaaatagagccacaggtgacagttaccctttggcattgac	2292
Db	2209	ACTGGCATTTGCATATCCCGCTCGCATTTA--CCAGACATAAATTTGGCTTTGCTGGTTCT	2265
Qy	2293	caaaagtaatggtctcacacgcctaaagctgacccgggtgagcagatacaaaatggtaatcga	2352
Db	2266	GATGGTGAGTTGATAC-----AAACAAACCTTATCTTTGATCAAGCAAGCTA--CAA	2316
Qy	2353	tgtgttattgagcaagtccctagcgtgacggttaacagcaccacaaasacacattaaagga	2412
Db	2317	GTGGCAATGTTTAAAGATTACCAACACTGGCATTTACGCAAGTGGTAAAGCCATCACAGG	2376
Qy	2413	tgtgtcccaacactgcttagcattgccaggtcccaagtgccgcaacatagcaactggggcaat	2472
Db	2377	CTGTCCCAACACTGCCTAGCATTCGCCGATCAAGTAGCCGCAACATAGAACTCGGCAAT	2436
Qy	2473	acaatcgaagaaaagacagaaaatccaaacgtccagcattgtagtgcgtaaatgcagggc	2532
Db	2437	ACAATCCAAAGCAAAAGACAAATTCCAAACGCTGCCAGCATTTAATGATATATTTAAATACAGGC	2496
Qy	2533	tttaacctaaaaataatggcgaagacaaaagactttgtctccactatgacactggtgac	2592
Db	2497	TTTTAACTTAAAAAATTAATACAAACCCATTGACTTTGTCTCCACTTATGACATTTGTTGAC	2556
Qy	2593	tttatcgatggcaatgccaccaccgccacagtaacttatgatgaagcgaacacaaacagt	2652
Db	2557	TTTTGCCAATGGCAATGCCACACCGCCACAGTAACCCATGATACCGCTTAACNAACCACT	2616
Qy	2653	aaagtggctgatgtgaaatggatgtagaagaaacacttgaaactgacaggcgataatggc	2712
Db	2617	AAAGTGGTATATGATGTGAATGTGGATGATACAAACCTTCAATCAACAGGCACTGATGAC	2676
Qy	2713	aagaaacaaacttggcgtcaaaacacatacaactgacgaagcaaacagtaactgaatgca	2772
Db	2677	AATAAAAAATTTGGCGTCAAAACACCAACTGAACAAACAAAGTGCTAATGGTAAATACA	2736
Qy	2773	actacat-----ttagtacgcagatgacacatgcctctgttaaagccagtgatc	2823
Db	2737	GCAACTTAACTTAATGTTAACTCTAGTGATGAAGATGCCCTTTGTTAACGCCAAAGACATC	2796
Qy	2824	gccggcaattcaaacacacctagccgaggaataatcaacacacaaagccagcacaacacc	2883
Db	2797	GCCGAAAAATCTTAAACACCTTAGCCACAGGAATTTACACACCACCAAGCGACAGCAGACAC	2856
Qy	2884	gccctacaaacttaccgtttaaaggttagatgaaatgataaggtctgatgacacacac	2943
Db	2857	GCCCTTACAAACCTTTACCGTTTAAAGGTAGATGAAATAATAATGTGATGACGCCAAC	2916
Qy	2944	gccatcacctgggttaaagaatggcaaatggttaaagtcacacaccttaaacctcaaaagt	3003
Db	2917	GCCATCACGGTGGGTCAAAAAGAACGCAAAATTAATCAAGTCAACACCTTAACACCTCAAGGT	2976
Qy	3004	aaaaacggtcttgattataaacacgacaaagatggtacgggttaacctttggcattaaacac	3063
Db	2977	GAAAAACGGTCTTAATATTTAAAAACCCACAAAAAATGGTACGGTTACCTTTGGCATTAACACC	3036
Qy	3064	caaacgggtcttaaacgcggcgagcaccactcttaaacacaatggcttgtctattaaa	3123
Db	3037	ACAAAGCGGTCTTAAAGCGGGCAAAAAGCAC--CTAAACGACGGTGGCTGTCTATTATAA	3093
Qy	3124	aacacgcgtagtaacgaacaaatccaagtcggtgatggcgtgaaagtttggcatggtt	3182
Db	3094	AACCCACCTGGTATGGGAACAAATCCAAAGTCGGTCTGATGGCTGAAGTTTGCCAAAGTTT	3153
Qy	3183	--taataatgggtgtgtaggtgctggcatgtatggcacaactgcacattaccagagatgaa	3240

Db 3154 AATAAATGTTGTTAGTGGCTTGGCATTTGATGGCAACATCGCATTTACCAGAGATGAA 3213  
|||||  
Qy 3241 attgctttactggagactaaaggctacttgaataaaagcaaacccacaccaaagac 3300  
|||||  
Db 3214 ATTGGCTTTACTGGACTAATGGCTCACTTGATAAAGCAAAACCCACCTTAAGCAAGAC 3273  
|||||  
Qy 3301 ggcattaacgcaggtgtaaaagattaccaaatccattcaatcagtgagattgcccataaac 3360  
|||||  
Db 3274 GGCATTAAACGAGTGGTAAAGATTACCAACATTTCAATCAGGTGAGATTGCCCAAAAC 3333  
|||||  
Qy 3361 agccatgatctgtgacagcgcaagatttatgatttaaaacccgaacttgaataaaa 3420  
|||||  
Db 3334 AGCCATGATGCTGTGACAGCGGCAAGATTATGATTAAACCCGAACCTTGAACCAAA 3393  
|||||  
Qy 3421 atcagcagttactgcataaacagcacaaaactcattcacgaattctcagtagcagatgaa 3480  
|||||  
Db 3394 ATCAGCAGTACTGCCAAACAGCACAAACACTATTACACGAATTTCTCAGTAGCAGATGAA 3453  
|||||  
Qy 3481 caaggtataactttacggttagtaacccttactccagtttatgacacctcaaaagacctct 3540  
|||||  
Db 3454 CAAGGTAAATAACTTTAGCGTTAGTAACCTTACTTCCAGTTATGACACCTCAAAAGACCTCT 3513  
|||||  
Qy 3541 gatctatccctttgcaggtgaaacggcattacacccaaggtataataaagggtgtggtg 3600  
|||||  
Db 3514 GATGTCATCACTTTTGCAGGTGAAACGGCATTTACCAACCAAGGTAAATAAGGTGTGGTG 3573  
|||||  
Qy 3601 cgtgtggcattgacaaaacaaaggcttaaccacgcctaaagctgacgctgggttaataat 3660  
|||||  
Db 3574 CGTGTGGCATTTACCAAAACCAAGGCTTAACCAACGCTTAAGCTGACCGTGGGTAAATAAT 3633  
|||||  
Qy 3661 aatggcaaggcattgtctattacacagccaaatgggtcacaataccatcacaggactaagc 3720  
|||||  
Db 3634 AATGGCAAGGCATTTGTCATTGACAGCCAAATGGTCAAAATACCATCACAGGACTAAGC 3693  
|||||  
Qy 3721 aacacttagctaatgttaccatgataaaggtagctacgcaccacagaacagggaat 3780  
|||||  
Db 3694 AACACTCTAGCTAATGTTTACCATGATAAAGGTAGCTAGCTACGCACACAGACAGGGCAAT 3753  
|||||  
Qy 3781 ataataaagacgaagacaaaacccgtgcgcagcattgttgatgtctaaagcgagcg 3840  
|||||  
Db 3754 ATAATCAAGACGAAGACAAAACCCGTCGCCAGCATTTGTGATGTCTAAGCGCAGGC 3813  
|||||  
Qy 3841 tttaactgcaaggcaatggtgaagcggttgactttgtctccacttatgacacgcgtcaac 3900  
|||||  
Db 3814 TTTAACTTGAAGCAATGTTGAAGCGGTTGACTTTTGTCTCCACTTATGACACCGTCAAC 3873  
|||||  
Qy 3901 ttgcgaatggcaataccacacgcctaaagtgacacctatgatgacacagaacacagat 3960  
|||||  
Db 3874 TTTGCCGATGGCAATGCCACCCGCTAAGGTGACCTTATGATGACACAGCAAAACCCAGT 3933  
|||||  
Qy 3961 aaagtgtctatgatgtcaatgtgatgtatacaaccatttgaagttaaaagataaaaaactt 4020  
|||||  
Db 3934 AAAGTGTCTATGATGTCATTTGATGATACACCATTTGAAGTTAAAGATAAAACACTT 3993  
|||||  
Qy 4021 ggcataaaacccacacattgaccagttactggcagaggtgctataataattgcccctaagc 4080  
|||||  
Db 3994 GGCCTAAACACCAACATTGACCACTGACCAAGGTGACAGAGGTGCTAATAAATTTGCCCTAAGC 4053  
|||||  
Qy 4081 aatcaagctactggcagctgcttgcaggccagtgatctatcgctgtctatctataaacacc 4140  
|||||  
Db 4054 AATCAAGCTACTGGCGATGGCGCTTGTCAAGGCCAGTATGATGTTGCTCATCTTAAACACC 4113  
|||||  
Qy 4141 ttatctggcgacatcccaactgccaaaggcgcaagcgaagcaaacactcagcagggctat 4200  
|||||  
Db 4114 TTATCTGGCGACATCCAAACTGCCAAAGGGCGCAAGCCAAACCACTCAGCAGGCTAT 4173  
|||||  
Qy 4201 gtgatgtgatgcaataagggtcatctatgacagttaccgataaacaagtaactcaagcc 4260  
|||||  
Db 4174 GTGGATGCTGATGGCAATAAGGTCTATGACAGTACCAGTACCAAGTACTATCAAGCC 4233  
|||||  
Qy 4261 aaaaatgatggcacagttgataaaaccaaagaagtttcccaaaagacaaactggtgcgcca 4320  
|||||

Db 4234 AAAAATGATGCGACAGTTTGATAAAACCAAGAAAGTTGCAAAAGACAAACTGGTGGCCCAA 4293  
|||||  
Qy 4321 gccaaaccccgagatggcacattgctcaaatgaatgtcaaatcagttcatlaacaaagaa 4380  
|||||  
Db 4294 GCCCAAAACCCAGATGGCACATTGGCTCAATGAATGAATGTCATTAACCAAGAA 4353  
|||||  
Qy 4381 caagtataatgatgccataaaaaagcaaggcatcaatgaagcaaacgcctttgtttaaagga 4440  
|||||  
Db 4354 CAAGTAATGATGCAATATAAAAGCAGGCATCAATGAAGCAACACGCCCTTTGTTAAAGGA 4413  
|||||  
Qy 4441 cttaaaaaagccgcttctgtataaaaaaaccaaaacccgagtaaacgttggttgattta 4500  
|||||  
Db 4414 CTTGAAAAGCCGCTTCTGTATAACAAAACCAACCCGCGCAGTAACCTGTGGGTGATTTA 4473  
|||||  
Qy 4501 aatccggttgcacaaacacgcgtacctttgcaggggatacaggcacacgcgtlaaaaaa 4560  
|||||  
Db 4474 AATCCGTTGCCAAACACCGCTGACCTTTGCAGGGGATACAGGCACACCGCTTAANAAA 4533  
|||||  
Qy 4561 ctggggagacatttgaccatcaaaaggtgggcaaacagacacacaataaagtaaacgcataat 4620  
|||||  
Db 4534 CTGGCGAGACTTTGACCATCAAAAGGTGGCAAAACAGACACCAATAAGCTAACCGATAAT 4593  
|||||  
Qy 4621 aacatcggtgtgtagcaggttactgagcttactgcttactgtcaaacitggccaaagacctaac 4680  
|||||  
Db 4594 AACATCGGTGTGGTAGCAGGTACTGTAGGTCTACCTGTCAAACTTTGCCAAAGACTAACCC 4653  
|||||  
Qy 4681 aatcttaacagcgttaatgacaggtggcaccacaaatgtgaaaggcatctcttttga 4740  
|||||  
Db 4654 AATCTTAACACCGTTAATGACAGGTGGCACCAAAATTTGATGACAAAGGCGTGTCTTTTGTGA 4713  
|||||  
Qy 4741 gacgcaaacggtcgaagccaaagcaaacacccctgtgctaaatggcctggacgtg 4800  
|||||  
Db 4714 GACTCAAGCGGTCAAGCCAAAGCAAAACACCCCTGTGTCTTAAGTGCCCAATGGGTGGACCTG 4773  
|||||  
Qy 4801 ggtggcaggttcacagtaaatgtgggcaaaagcacaaaagataccgacgcgtgccaatga 4860  
|||||  
Db 4774 GGTGGCAGGTCTACAGTAATGTGGGCAAGGCAACAAAGATACCGACGCTGCCAATGTA 4833  
|||||  
Qy 4861 caacagttaaacaggttacgcaactgttgggtctt-----tggttaatgataacgct 4911  
|||||  
Db 4834 CAACAGTTAAACGAAGTACGCAACTTTGTTGGTCTTGGTAAATGTGGTTAATGATPAACGCT 4893  
|||||  
Qy 4912 gacggcaatcaggttaaacattgcccagacatcaaaaaagaccccaaatcaggttcatctct 4971  
|||||  
Db 4894 GACGGCAATCAGGTAAACATTCGGACATCAAAAAGACCCCAAAATTCAGTTTCATCATCT 4953  
|||||  
Qy 4972 aaccgacgttcatacaagcagcgactgtactggcgtaaaagtaataacagataccgaa 5031  
|||||  
Db 4954 AACCGCACTGTCTATCAAAAGCAGCGACGGTACTTTGGCGGTAAAGGTAATAACGATACCGAA 5013  
|||||  
Qy 5032 aaacttgccactgggtgtgtacaagtgggcggtgataaagcgcgaacgcgttaacggcgat 5091  
|||||  
Db 5014 AAATTCGCACCTGGTGATATCAAGTGGCGTGGATTAAGACGGCAACGCTTAACGGCGAT 5073  
|||||  
Qy 5092 ttaagcaatgttgggtgtcaaaaccccaaaagatggcagcaaaaaagccctgtctgcccact 5151  
|||||  
Db 5074 TTAAGCAATGTTGGGTCAAAACCCCAAAAGATGGCAGCAAAAAGCCCTGCTCGCCACT 5133  
|||||  
Qy 5152 tataacgcgagctcagacaccaactatgtgaccaacaaaccccgagaagccattgacaga 5211  
|||||  
Db 5134 TATAACCGCGCAGGTCTAGACCAACTATTTTGACCAACAAACCCCGCAGAGCCATTGACAGA 5193  
|||||  
Qy 5212 ataatgaacaaggtatccgctcttccatgtcaacgtatggcaatcaagagcgtgtgta 5271  
|||||  
Db 5194 ATAATGAACAAGGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGACGCTGTGGTA 5253  
|||||  
Qy 5272 caagggcgtaacggcattgactcaagtgcctcaggcaagcactcagttggcgatagggttc 5331  
|||||  
Db 5254 CAAGGGCGTAACGGCTTGAAGTCAAGTGCCTCAGCAAGCACTCAGTGGCGATAGGTTTC 5313  
|||||  
Qy 5332 caggccaaaggcagatggtgaagcgccgctgtcccataggcagacaaaccccaagcaggcaac 5391  
|||||  
Db 5314 CAGGCCAAGGCAGATGGTGAAGCCGCCCTTGCCATAGGCAGACAAACCAAGCAGGCAAC 5373  
|||||





Qy 943 tctacggatataatttttccatctggttaagtagtaataataataatagcagtatcagggcgtaaa 1002  
Dy 1490 AAGCGGGTCCACTTTCCATTTGGTAGTAA-----CTCTATTCAAAACGTA 1534  
Qy 1003 atcatcaatgctgggtggttctcgggataaccggtccggtcaatggtgagcagcttaaa 1062  
Dy 1535 ATCATCAATGTCGGTGCAGGTTTAATAAAACCGATCGGTCATGTGGCAGCAGTAGAA 1594  
Qy 1063 ttggtggagggaactgggcta-----atcgtaaaattacttttaagggtgagtgacaataat 1119  
Dy 1595 GCGGTGGTGAAGTGGCTAAGGAGCGTAGAATTACTTTTCAGGGTGATG-----ATAAC 1648  
Qy 1120 agcaatagcgtagagaaggggtttgggcaactttaaactttaaaggtgagtcacagacc 1179  
Dy 1649 AGTACTACGCTAAAGATAGGTTTGGATAAATACTTTAACTATTAAGGTGGTCCAGAGACC 1708  
Qy 1180 aacgattaaacggagctaaactcgggtggtgaacaaatgga-----tggtctgaaa 1233  
Dy 1709 AACCATTAAACCGATATAATATCGGTGTGTAAGAGGCTGATAATAGTGGTCTGAAA 1768  
Qy 1234 gttaaactgtcgaagagctgactggattgaccaggtgtctcc-----gct 1278  
Dy 1769 GTTAACTTGTCTAAACTTTTAAACAATCTTACTGAGGTGAATACAACTTAAATGCC 1828  
Qy 1279 accaacaataatcccggttagtaataccaacaacaacacgagcgcagctacaagcggtggt 1338  
Dy 1829 ACAACCCAGTTAAGGTAGGTAGTAGTAGTACTACAGCTGAATTTATGAGTGATAGT 1888  
Qy 1339 ttgacccctttagccca-----atacagggtacaaaaacagataaaacgctctacagcatt 1392  
Dy 1889 TTAACCTTTACCCAGCCCAATACAGGCGAGTCAAGCAGCAACAAACCGTCTATGGCGTT 1948  
Qy 1393 gatgattgaagtttactaataatagtagtaataatgtagcaactaaagggtactactcgtatt 1452  
Dy 1949 AATGGGTGAAGTTTACTTAATAATGCAGAAACACAGCAGCAATCGGCATCTCGTATT 2008  
Qy 1453 accaaaaaagaataatggttttgcgttactaatatgtaggagttgatgaaagcaaaccttat 1512  
Dy 2009 ACCAGAGATAAATTTGGCTTTCGAGAGTGTGTGATGTGTGATGATAAATAAACAAGCACCATAT 2068  
Qy 1513 ctgacacgaaaagctaaagttggcaacagcaccctaaacagtggttagcttaacttt 1572  
Dy 2069 TTGATAAATAAACCACTTAATAGTGGTAGTGTTCGAATTTACCATAGACAATGGCATTTGAT 2128  
Qy 1573 aatacaccactggttaataaacaatacccaagtcggtgctgtaagtcatttaaatttggccaca 1632  
Dy 2129 GCAGGTATAAAGATCAGTAATTTGCCAAAGGTACAGTGTCTAACGATCCGGTTACC 2188  
Qy 1633 gtcgctaataatgltgcaaatatctctcagcaacagtcggcactgctcgtattaccogaagag 1692  
Dy 2189 ATCGAACAGCTCAAGCGCCCAAGCCTACTTTAAACGAGCGGCTGGCATCAGTGTCTACA 2248  
Qy 1693 aaaattggttttgcgttactaataatgtaggagttgatgaacaagcaccataatttggataaa 1752  
Dy 2249 CCTACTGAATATCAGTTGATGCTTAAGAGTGGCAATGTTTACCGCCCCAACCTTACAACA-- 2306  
Qy 1753 gaagacttaagtggttcggtgttgaaattaccacagatggtgtattaatgctggttaatt 1812  
Dy 2307 -----TTGGCGTGAAACACCCAGGCTTAACAGTGATGCGCAGTGTGATAAATTTAGT 2359  
Qy 1813 cacaagattaccgacttactaataatggtatagcaataaccgtagtgggttaccatcaaacag 1872  
Dy 2360 GTTAAGGGTAGTGTACCAACAATAGCTTAGTTACCGCCGAACAATTTGGCAAGCTATCTA 2419  
Qy 1873 ctcaaaagcgaagcctactttaaagcagcgagctggtcatcagatttaataagtaataac 1932  
Dy 2420 AATGAGTCAATCGAACGGCTTGACAGTGTCTACAAAGCTTTTACCGTTAAGAGAAGAAC 2479  
Qy 1933 ggggatctagttgatagtggtggaataattaccaccccaacttaaacattagcgtgaaa 1992  
Dy 2480 GATGATGAGC-----CCNACGGCTATCACCGTGCT 2509

Qy 1993 accactaagcttaacagtaattggcaccagtggttaataataataatttagttagtaagtct 2052  
Dy 2510 AAAGATACGACAAATAATCGCGGCAGTCAGCATCTTTAAAACTCAAAAGTAAAAACGGT 2569  
Qy 2053 catgataacaatagcttagttaccgccaagaatttggcagactatcctaataaagtcaat 2112  
Dy 2570 CTACGGTTGCTACCAAAAAGATGGTAGGGTTACCTTTTGGGCTTAGCCAGATAGCGGT 2629  
Qy 2113 gaaacgggtgcagctgctctaccaagcttttaaagtcctcaaaacgggtgataaagcaaac 2172  
Dy 2630 CTGACCATTTGCAAAAGCACCCTAAACACGATGGCTTGTACTGTTTAAAGATACCAACGAA 2689  
Qy 2173 gccatccacgctgggtaaagatacaaaacggcgaagacctcaaacaccttaaaactcaaaagt 2232  
Dy 2690 CAAATCCAAAGTCGGTCTAATGGCATTAATTTTACTAATGTGAATGGTAGTAATCCAGT 2749  
Qy 2233 gaaacgggtgtaataattacgaccaatagagccacagctacagttacgtttggcattgac 2292  
Dy 2750 ACTGGCATTTGCAATACCGCTCGCATTTA---CCAGAGATAAATTTGGCTTTTGTGGTCT 2806  
Qy 2293 caaagtaattggtctcaccacgcttaagctgacgctgggtgagcgtacaaaatggttaatcga 2352  
Dy 2807 GATGTGCAGTTGATAC-----AACAAACCTTTATCTTGATCAAGACAAGCTA--CAA 2857  
Qy 2353 ttggttattgagcaggtccctagcgtgacggtaacagcacaacaaacatcattaaagga 2412  
Dy 2858 GTTGGCAATTTAAGATTTACCAACACTGGCATTAACGCGAGGTGGTAAAGCCATCAGGG 2917  
Qy 2413 ttgtcccaacacgtcctagcattgcccagttccagtcgaagtcgcccgaacatgacgtggcaat 2472  
Dy 2918 CTGTCCCAACACCTGCCTAGCATTCGCGATCAAAAGTAGCCGCAACATAGAACTGGGCAAT 2977  
Qy 2473 acaatcgagaagaaaaagacaaatcccaacgctgcagcattgtagtgctaaatcgagc 2532  
Dy 2978 ACAATCCAAGACAAAGACAAATCCAACGCTGCCAGCATTAATGATATATTAATACAGC 3037  
Qy 2533 tttaacctaaaaataatggaagacaaagacttgtctccacttgaacactgttgcac 2592  
Dy 3038 TTTAAACCTTAAAAATTAATAACACCCCATTTGACTTTGTCTCCACTTATGACATTTTGAC 3097  
Qy 2593 ttatcgtatgcaatgcccacccgcccagctaaacttatgatgaagcacaatacaaacaggt 2652  
Dy 3098 TTTGCCAATGGCAATGCCACCCACCCACAGTAACCCATGATACCCGTAAACAAAACCCAGT 3157  
Qy 2653 aaagtggcgtatgtagtgaattggtgagaaaaacattgaactgacagggcgataatggc 2712  
Dy 3158 AAAGTGGTATATGATGTGAATGTGGATGATACAACTTCACTAACAGGCATGATGAC 3217  
Qy 2713 aagaacaacttggtcgtcaaaacacatcaaaactgacccgaacaaagactaattggtaatgca 2772  
Dy 3218 AATAAAAACTTGGCGTCAAAACCCACCAAACTGAACAAAAGTGTCTAATGGTAATACA 3277  
Qy 2773 actaat-----ttagtagcagcagatgacccgttgccttgggttaaagcagtagtatc 2823  
Dy 3278 GCAACTACTTTTAATGTTAACTCTTAGTGAAGATGCGCTTGTAAACGCCAAAGACATC 3337  
Qy 2824 gccggcaacttaaacacccctagcaggaataattcacaccacaaagggcacagcaaacacc 2883  
Dy 3338 GCCGAAATCTAACACCCCTAGCCAAAGGAATTCACACCACCAAGGCACACAGCACACC 3397  
Qy 2884 gccctacaaaacctttaccgttaaaaaggttagatgataaataagataagagctgatgacaccaac 2943  
Dy 3398 GCCCTACAAAACCTTTACCGTTAAAAAGGTAGATGAAAATAATATGCTGATGATGACGCCAAC 3457  
Qy 2944 gccatccggtgggttaagatggcaagtggttaagttcaacaccttaaaactcaaaagt 3003  
Dy 3458 GCCATCACCGTGGGTCAAAAGAACGCAATAATCAAGTCAACACCCCTAACACTCAAAAGGT 3517  
Qy 3004 aaaaacgctcttgattataaacacgacaaagatggttagcgtttacgtttggttaaacacc 3063  
Dy 3518 GAAACGCTCTTAATATTAAACCCACCAAAATGTTACGGTTACCTTTGGCATTAACACC 3577  
Qy 3064 caaagcgggtcttaagacggcgagcaccactcttaaacacaatggtttgtctataaa 3123



Qy 5272 caagggcgtaacggcattgaactcaagtgcctcctcagggcaagcaactcagtggtgagataggtttc 5331  
 Db 5795 CAAGGGCGTAACGGCATTGACTCAAGTGCCCTCAGGCAAGCACTCAGTGGCGCATAGGTTTC 5854  
 Qy 5332 caggccaagggcagatggtgaagccgcgtgtccctataggcagacaaacccagcaggaac 5391  
 Db 5855 CAGGCCAAGGCAGATGTTGAAGCCGCCGTGTCCATAGGCAGACAAACCAAGCAGGCCAAC 5914  
 Qy 5392 caatccatgcgcattgcgtgataacgcacaaagccagggcgatcaatccatgcgcattgcgt 5451  
 Db 5915 CAATCCATGCCCATCGGTGATGAACGCAACAGCCAGCGGCGATCAATCCATGCCATCGT 5974  
 Qy 5452 acaggcaatggtgtagcaggtgaacactctgtgtccatcgccagcccaagcaactgttaag 5511  
 Db 5975 ACAGGCAATGTTGGTAGCAGGTGAACACTCTGTTGCCATCGCGACCCCAAGCACTGTTAAG 6034  
 Qy 5512 gctgataacagttacagtggtggtgataaacaacacagttaccgatgccaactcaaacagat 5571  
 Db 6035 GCTGATACACTTACAGTGTGGTGAATAAACAAACAGCTTTACCGATGCGACATCAACCGAT 6094  
 Qy 5572 gtcttgggtggaataacatcacctgacccggaagaaactcgtgtgctttaggttca 5631  
 Db 6095 GTCTTTGGTGTGGCAATAACATCACCGTGACCGAAAGTAACCTCGTGTGCCCTTAGGTTCA 6154  
 Qy 5632 aactctgcattcagtcagggcacacagcagcagcagcagcagcagcagcagcagcagcagc 5691  
 Db 6155 AACTCTGCATCAGTGCAGGCACACACGACGACGACACCAAGCAAAATCTGACGGCAC 6214  
 Qy 5692 gcaggtcaaacacacacagcaggtgacacaggtacaggttaagccttgcgtggacaacg 5751  
 Db 6215 GCAGGTCAACACCAACACAGCAGGTGCAACCGGTACGGTTAAAGGCTTTGCTGGCAAAACG 6274  
 Qy 5752 gcggtggtgcggtctcgtggtggtgcctcaggtgctgaacccggtatcccaaaatgtgca 5811  
 Db 6275 CGGTTGTTGCGGCTCTCGTGGTGCCTCAGGTGCTGAACGCGGTATCCAAATGTGGCA 6334  
 Qy 5812 gcaggtgaggtcagtgccacagcagcagcagcagcagcagcagcagcagcagcagcagc 5871  
 Db 6335 GCAGGTGAGGTGAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 6394  
 Qy 5872 acccaagcattgcgaacgcgaacaaatgagcttgacctgtatcccaacaaacgaacaaat 5931  
 Db 6395 ACCCAAGCATTGCACAGCAACCAATGAGCTTGACCATCGTATCCACCAAAACGAAAT 6454  
 Qy 5932 aaagccaatgcagggatttcacagcagcagcagcagcagcagcagcagcagcagcagcagc 5991  
 Db 6455 AAGCCAAATGCAGGATTTTCATCAGCGATGCGGTGCGGTCCATGCCACAAGCCTACATT 6514  
 Qy 5992 cctggcagatccatggttacccgggggtattgcccacccacacacaggtcgaaggtgcgggca 6051  
 Db 6515 CCTGGCAGATCCATGTTTACCGGGGGTATTGCCACCCCAACAGCTCAAGGTGCGGTGGCA 6574  
 Qy 6052 gtggactgtcgaagctgcggataatggtcaatgggtattataaatacaatggttcagcc 6111  
 Db 6575 GTGGACTGTGCAAGCTGCGGATAATGGTCAATGGGTATTAATTAATTAATTAATTAATTA 6634  
 Qy 6112 gatcccaagccatgtagggcgccaggtggtgagcaggttttcattt 6159  
 Db 6635 GATACCCCAAGCCATGTAGGGGGCGCAGTTGTCAGGTTTTCATT 6682

RESULT 4  
 AR040716 6973 bp DNA PAT 29-SEP-1999  
 LOCUS Sequence 1 from patent US 5808024.  
 DEFINITION AR040716  
 ACCESSION AR040716  
 VERSION AR040716.1 GI:5960079  
 KEYWORDS Unknown.  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 6973)

AUTHORS Sasaki, K., Harkness, R.E., Loosmore, S.M. and Klein, M.H.  
 TITLE Nucleic acids encoding high molecular weight major outer membrane protein of moraxella  
 JOURNAL Patent: US 5808024-A 1 15-SEP-1998;  
 FEATURES Location/Qualifiers  
 source 1..6973  
 BASE COUNT 2265 a 1555 c 1533 g 1620 t  
 ORIGIN  
 Query Match 66.4%; Score 4090.6; DB 9; Length 6973;  
 Best Local Similarity 81.3%; Pred. No. 0;  
 Matches 5063; Conservative 0; Mismatches 1009; Indels 157; Gaps 21;  
 Qy 1 atgaatcacatctataaagtcatcttttaacaaagccacagggcacatttatggccgtggcg 60  
 Db 542 ATGAATCACATCTATAAAGTCATCTTTAACAAAGCCACAGGCACATTTATGCGACGTGGCA 601  
 Qy 61 gaatatgcacaaatccacacagttac-ggggggggggtagctgtctacagggcaagttggcag 119  
 Db 602 GAGTACGCCAAATCCACAGCAGCGGGGGGGTGTGTGTACAGGGCAAGTTGGCAG 661  
 Qy 120 tgtacgcaactcaagctttgcccgtattgcgcgcgtcgtcgtcctcgtgagtcggtgcag 179  
 Db 662 TGTATGCACTCTGAGCTTTGCCCGGTATTGCCGCGCTCGCTGCTCGTATCGGTGCAAC 721  
 Qy 180 gctcaatggcagtgcttattgtccaacaaattactacca---agatcgaaattgggtcaaac 236  
 Db 722 GCTCAGTGGCAGTGCTTATGTCTCAAAAAAAGATACCAAAACATATCGCAATTTGGTGAACA 781  
 Qy 237 aaacaagataaacacacgcgtgaaggcgtgcccctagcagcaggtgaagcatccattgc 296  
 Db 782 AAACAGCAAGACGCTC-----AGGCACCTGCCAAGCGGACGGTGTATCGAGCCATTGC 835  
 Qy 297 ttttggtgagttcttaaggcacaaggtctcaagctattgctatcggttagtgtcaaac 356  
 Db 836 TATTGGTGAAATGTCTAACGACACAGGGCGGTCAAGCCATFCGCCATCGGTAGTAGTAATAA 895  
 Qy 357 agatcctaataatggttagtaattggttaagttggttcctcacgcaaggttaacgagtc 416  
 Db 896 AACTGTCAATGGAAGCAGTTTGGATAAGATAGTACCGCATGCTACGGGTCAAGAGTCCAT 955  
 Qy 417 cgcacatgggtggtgagtgattgggtgaggtgagtcctcgtcgtcgtcgtgagtgatga 476  
 Db 956 CGCCATCGGTGAGTGTAAAGGCTAGTGTGATGCTGATGCTGATGCTGATGCTGATGATGA 1015  
 Qy 477 cttatatttgctgaagaatcttgatctgaagaa-----tgaatttcacaa 521  
 Db 1016 CTTACATTTGCTTGATCAGCATGTTGTAATCCTTAAACATCCGAAAGGTACTCTGATTAAACGA 1075  
 Qy 522 acttattcattgcccataaataataataataataataataataataataataataataataata 581  
 Db 1076 TCTTATTAAAGCCCATGTCAGTATTAAAGAAATACGAAGCTCAAGAGGATAATGATGTAA 1135  
 Qy 582 atatacgcacaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 641  
 Db 1136 ATATAGCGCACACACCGAAGCGGACAGCCAGTACTGTCAGTGGGAGCCATGTCATATGC 1195  
 Qy 642 acaggtcatttttcccaagcctttggtacatacacaacagcagcagcagcagcagcagcagc 701  
 Db 1196 ACAGGGTCATTTTCCACGCGCTTTGGTACACGGGCAACAGTAAAGGTGCCTATTCCTT 1255  
 Qy 702 ggcagtaggtcttgcgcgcacaaagccacaaacaaactctcaatcgctgttggttccaaatgc 761  
 Db 1256 GGCAGTGGGTCTTGCGCCACAGCGGCGCAATCTACATCGCTATTTGGTCTTGATGTC 1315  
 Qy 762 aaaagcacaacgcttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 821  
 Db 1316 AACATCTAGCTGTTGGGAGCGATAGCCCTTGGTGCAGGTACTCGTGTCTGCTACAGG 1375  
 Qy 822 aggcgttgcctaggttttgggttcagatccttgatagggaataataacagagcagcag 881



QY 3003 taaaaacgggtcttgatattataaaacccgacaaagatgggtacggttaccttggcattaacac 3062  
Db 3518 TGAACACGGTCTTAATATTAACCGCAAAATATGGTACGGTTTACCTTTGGCATTAACAC 3577  
QY 3063 ccaagcgggtcttaaacccgagcagcacacactctataaacaacaaatggcttgcctattaa 3122  
Db 3578 CACAAGCGGTCTTAAGCGCGCAAAAGCACG- --CTAAACGACCGGTGGCTTGTCTATTAA 3634  
QY 3123 aaacacgcgttagtaacgaacaaatcccaagtcgggtgctgtagggcgtgaagttggcattgg 3182  
Db 3635 AAACCCCACTGGTAGCGCAACAAATCCAAGTCGGTGCCTGATGGCGCTGAAGTTTGCCAAAGT 3694  
QY 3183 ---taataagtggtttagtgctggtgattgagtcagacaaactgcgattaccagagatga 3239  
Db 3695 TAATAAATAGTGTGTAGGTGCTGGCATTTGATGGCACAACCTGCATTAACGAGATGA 3754  
QY 3240 aattgggttactgggactaatggctcacttgataaagcaaacccacacctaagaaga 3299  
Db 3755 AATTGGCTTTACTTGGGACTAATGGCTCCTCTGTATTAAGCAAAACCCCACTTAAGCAAGA 3814  
QY 3300 cggcattaacgcaggtggtataaaagattacaaacattcaatcaggtgagattggcaaaaa 3359  
Db 3815 CGGCATTAAACGAGTGGTAAAGAGATTACCAACATTCAATCAGGTGAGATTGCCCAAA 3874  
QY 3360 cagccatgactgtagcagcggcgaagattattgatttaaaacccgaacttgaaaaataa 3419  
Db 3875 CAGCCATGATGCTGTGACAGCGCGCAAGATTATGATTTAAAAAACCGCAACTTGAACAA 3934  
QY 3420 aatcagcagtagtgcgaacacagcacaaaactcaatcacgaattctcagtagcagatga 3479  
Db 3935 AATCAGCAGTACTGCCAAACAGCACAACACTCATTACAGCAATCTCAGTAGCAGATGA 3994  
QY 3480 acaaggtaataacttactcggtagtaaccttactccagttatgacacctcaagaacctc 3539  
Db 3995 ACAAGGTAATAAATTTACGGTTAGTAACCTTACTCCAGTTATGACACTCAAAAGACCTC 4054  
QY 3540 tgatgtacacacttgcaggtgaaacgcgattaccacaaagtaataaagtggtggt 3599  
Db 4055 TGATGTATCATCACCCTTGCAGGTGAAACGCGCATTTACCACCAGGTAATAAAGGTGGT 4114  
QY 3600 gcgtgtggcattgacaaacaaaggcttaaccacgcctaaagctgacgcgtgggtaataa 3659  
Db 4115 GCGTGTGGCATTCACCAACCAAGGCTTAACACGCTTAAGCTGACCGTGGGTATATA 4174  
QY 3660 taatggcaaggcattgtcattaacagccaaaatgggtcaaaaataccatcacaggactaag 3719  
Db 4175 TAATGGCAAAAGGCATTTGTCATTGACGCCAAATATGGTCAAAATACCATCACAGGACTAAG 4234  
QY 3720 caacactcagctaattgttaccatgataaaggtagcgtacgcacacagaaacaggcaaa 3779  
Db 4235 CAACACTAGCTAGCTAATGTTACCAATGATAAAGGTAGCGTAGCCACACAGAACAGGGCAA 4294  
QY 3780 tataatcaaacgacgaacaaacccgtgcgcgcagcattgttgcgtgataacgcgcag 3839  
Db 4295 TATTAATCAAGACCAAGCAAAACCCGTGCGCCAGCATTTGTTGATGTGCTTAAGCGCAG 4354  
QY 3840 ctttaacttgcgaaggcaatgggtgaagcgggtgacttgcctccacttatgacacccgtcaa 3899  
Db 4355 CTTTAACCTTGAAGGCATGTGTGAAGCGGTTGACTTTGCTCCACTTATGACACCGCTCAA 4414  
QY 3900 ctttgcgaatggcaataaccacacgcgttaagtgacctatgatacacaagcaaaacag 3959  
Db 4415 CTTTGGCCGATGGCAATGGCCACACCGCTTAAGGTGACCTTATGATGACAAACCAAAACCCAG 4474  
QY 3960 taaagtggtctctatgattcaatgtggaatgatacaacacctgaagttaaagataaaaaact 4019  
Db 4475 TAAAGTGGTCTATGATGTCATGTGGATGATACAACCATTTCAAGTTAAAGATAAAAAACT 4534  
QY 4020 tggcgtataaaacaccacaaattgaccagtagcagcaggtgcttaataaatttgccttaag 4079  
Db 4535 TGGCGTAAAAACCAACCATTTGACAGTACTTGGCAGAGGTGCTAATAAATTTGCCCTAAG 4594  
QY 4080 caatcagctactggtgcgtggtgtcaaggccagtgatagcttgcctatcaaacac 4139

Db 4595 CAATCAAGCTACTGGCGATGCGTGTCAAGGCCAGTGATATCGTTGCTCATTAACAC 4654  
QY 4140 cttatcttggcagacatccaaactgcaaaaggggcgaagccaaagcgaacaaactcagcaggcta 4199  
Db 4655 CTTATCTTGGCGACATCCAACTGCCAAAGGGCAAGCCAAAGCAACTCAGCAGGCTA 4714  
QY 4200 tgtgtagtgcgtggaataaagggtcactatgacagttaccgatacaaaagtactatcaagc 4259  
Db 4715 TGTGGATGCTGATGCAATAAAGGTCTATGACAGTACCGATACAAAGTACTATCAAGC 4774  
QY 4260 caaaatgatgcagagttgataaaacaaagaagttgcgaagcaaaactggctgcacca 4319  
Db 4775 CAAAATGATGGCACAGTTGATAAAACCAAGAGTTGCCAAAGCAAACTGGTGCGCCA 4834  
QY 4320 agccaaaccccaagatggcacattggctcaaatgaatgcataatcagtcattaaacaaaga 4379  
Db 4835 AGCCCAAAACCCAGATGGCAGCATTTGGCTCAATGAATGCAATCAGTCATTAAACAAGA 4894  
QY 4380 acaagtaaatgatgccaaataaaagcaaggcatcaatgaagaacaacgcttggtaaagg 4439  
Db 4895 ACAAGTAAATGATGCAATAAAGCAAGGATCAATGAAGCAACACGCTTTGTTAAAGG 4954  
QY 4440 acttgaaaaacgcgtctctgataaaacaaacaaacccgagtaactgtgggtgattt 4499  
Db 4955 ACTTGAAAAAGCCGCTTCTGATAACAAACCAAAACCCGACGTAACGTGGGTGATTT 5014  
QY 4500 aatlgccgttgcgaacacacccgtgaccttggcaggggatacagcgaacacgcttaaaaa 4559  
Db 5015 AAATGCCGTTGCCCAACACCGCTGACCTTTGCAGGGGATACAGCACAACCGCTTAANA 5074  
QY 4560 actggggagactttgacctataaagggtgggcaaacagacaccaataaagctaaacgataa 4619  
Db 5075 ACTGGCGAGACTTTGACCATCAAGGTGGGCAACAGACACCAATAAGCTAACCGATAA 5134  
QY 4620 taacatcggtgtgtagcagtagcttgcctcactgcactcaaacctgcaaacgactaac 4679  
Db 5135 TAACATCGGTGTGGTAGCAGGTACTGTGGCTTCACTGTCAAACTTGCCAAGACCTAAC 5194  
QY 4680 caacttaacagcgttaatgcaggtggccacaaattgataaaaagcactctctttgt 4739  
Db 5195 CAATCTTAACAGCGTTAATGCGAGTGGCCACCAAAATTCATGACAAAGCGGTGCTTTGT 5254  
QY 4740 agacgcaaacggttcaagcgaagcaaacacccctgtgctaagtgcattgggtggaact 4799  
Db 5255 AGACTCAAGCGTCAAGCAAAAGCAAAACACCCCTGTGCTAAGTGCCAATGGGCTGAGCT 5314  
QY 4800 ggttggcaagtcactcagtaagtgtggcgaaggcacaagaataccgacgctgccaatgt 4859  
Db 5315 GGGTGGCAAGGTCTACGTAATGTGGGCAAAAGGCAAAAGATACCGACGCTGCCAATGT 5374  
QY 4860 acaacagttaaacgaagtacgcaacttgttgggtct-----tggtataatgaacgc 4910  
Db 5375 ACAACAGTTAAACGAAGTACGCAACTTCTTGGTCTTGGTAACTGCTTAATGATAACGC 5434  
QY 4911 tgaaggcaactcaggttaacatgtgcgacatacaaaaagaccacaaattcaggttcatc 4970  
Db 5435 TGACGGCAATCAGSTAAACATTTGCCGACATCAAAAAGACCCAAATTCAGGTTTCATC 5494  
QY 4971 taacgcactgtcatcaaacgagcagcagcttacttggcgttaagtaataacgataccga 5030  
Db 5495 TAACGCGACTGTCTCAAAAGCAGGCGGCTACTTGGCGGTAAAGGTAAATACGATACCGA 5554  
QY 5031 aaaacttccactggtggttacaagtggcgtggaataaaagcgaacgcgttaacgga 5090  
Db 5555 AAAACTTCCACTGCTGCTGATACAAAGTGGCGTGGATAAAGACGCAACGCTAACGGCA 5614  
QY 5091 ttaagcaatgttgggtcgaaccccaaaaagatggcgcaaaaaagccctgcgtcgcaac 5150  
Db 5615 TTTAAGCAATGTTTGGGTCAAAACCCAAAGATGGCAGCAAAAAGCCCTGCTGCCAC 5674  
QY 5151 ttataacgcgcaggttcagacaaactatgtaccaaacaccccgagagccattgaacag 5210



5675 TTATAACGCCGAGGTGACACCAACTATTGACCAACAACCCCGCAGAAGCCATTGACAG 5734  
QY 5211 ataatgaacaaaggtatccgtcttccatgtcaacgatggcaatcaagagcctgtggt 5270  
Db 5735 AATAAATGAACAAAGGTATCCGCTTCTTCATGTCAACGATGGCAATCAAGACCCCTGTGGT 5794  
QY 5271 acaaggcgtaacggcattgactcaagtgccctcaggcgaagcactcagtg99cgaagggttt 5330  
Db 5795 ACNAGGCGGTACGGCATTGACTCAAGTGCCCTCAGGCAAGCACTCAGTGGCGATAGGTTT 5854  
QY 5331 ccaggccaaggcagatggtgaagcgcgcttccatagcagacaaaccccaagcaggcaa 5390  
Db 5855 CCAGGCCAAGGCAGATGGTGAAGCGCGCTTGCCATATGCGACAGCAACCCCAAGCAGGCAA 5914  
QY 5391 ccaatccatgccatcggtgataaacgacaaagccacgagcggcgatcaatccatcgccatcg 5450  
Db 5915 CCAATCCATGCCCATCGGTGATTAACGACCAACGCCACGGCGGATCAATCGCATCGCATCGG 5974  
QY 5451 tacaggcaatggtgagcaggtaagcactctggtgccatcgcgcgacccaagcactgttaa 5510  
Db 5975 TACAGGCAATGTGATGACAGGTAAAGCACTCTGGTGGCATTCGGCGACCCCAAGCACTGTAA 6034  
QY 5511 ggcgtataacagttacagttggtggttaataaacaccggtttaccgatgcccactcaaacoga 5570  
Db 6035 GGCTGTATAACAGTTTACAGTGTGGGTATTAACACACCAAGTTTACCGATGCCACTCAAAACGA 6094  
QY 5571 tgtcttgggtgggcaataaacatccgtgacccgaagtaactcgtgctttaggttc 5630  
Db 6095 TGCTTTGGTGGGCAATTAACATACCGTGAACGCAAGCAAGTAACTCGGTGGCTTAGGTTTC 6154  
QY 5631 aaactctgccaatgacagcagcacacgagcagcagcagcagcagcagcagcagcagcagc 5690  
Db 6155 AAATCTGCCATCATAGTCAGGCGACACACGAGGACACACAGGCAACCAAAATCTCACGCCAC 6214  
QY 5691 aqcggttacaaccaccacagcaggtgcccacaggttaaggttaaggttctgctggacaac 5750  
Db 6215 AGCAGGTACACACCACACAGCAGGTGCAACGGGTACCGGTAAAGGCTTTGCTGGACAAAC 6274  
QY 5751 ggcggttgggtgctcgtggtgctcaggtgctgagcgcgctgacccaaatgtggtgc 5810  
Db 6275 GCGCGTGTGGTGCGGTCCTGGTGGGTGCGCTCAGTGTGCTGACGCGGTATCCAAATGTGGC 6334  
QY 5811 aqcggttgggttgcagtcaccacagcagcagcagcagcagcagcagcagcagcagcagc 5870  
Db 6335 ACCAGGTGAGGTCACTGCCACAGCAGCAGTCCGCTCAATGGTATGCGCAGTGTACAAAGC 6394  
QY 5871 caccacaagcattgccaacgaacaaatgagcttgaccatcgatccaccacaaacgaacaaa 5930  
Db 6395 CACCCAAAGCATTGCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 6454  
QY 5931 taaagccaatgcaggatgttcacagcagatggcgatggcgatggcgatggcgatggcgatgg 5990  
Db 6455 TAAGGCCAATGCAGGATTTATCAGCGATGGCGATGGCGTCCATGGCCACAAAGCCTACAT 6514  
QY 5991 tctgacagatccatggttacgggggtattgcccacacacagcgtcaggtgcggtggtgc 6050  
Db 6515 TCTGCGCATCATCATGTTTACCGGGGTATTCGCCACCCACACAGCGTCAAGGTGCGGTGCG 6574  
QY 6051 agtgggactgtcgaagctgctcggaataatggtcaatgggttatttaaaatcaatggttcagc 6110  
Db 6575 AGTGGGNCGTGCAAGCTGTCGATTAATGTCATATGGGTATTTAAATCAATGTTGTCAGC 6634  
QY 6111 cgataccaaaggccatgtagggcgagcttggttgggttgggttgggttgggttgggttgggt 6159  
Db 6635 CGATACCAAGGCCATGTAGGGGCGGAGTTGGTGACAGTTTTCACATTTT 6683

RESULT 5  
AX079913 LOCUS AX079913 6973 bp DNA PAT 22-FEB-2001  
DEFINITION Sequence 1 from Patent WO0107619.  
ACCESSION AX079913  
VERSION AX079913.1 GI:13159436

KEYWORDS Moraxella catarrhalis.  
SOURCE Moraxella catarrhalis  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;  
Moraxella.  
REFERENCE 1 (bases 1 to 6973)  
AUTHORS Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.  
TITLE Recombinant high molecular weight major outer membrane protein of  
Moraxella  
JOURNAL Patent: WO 0107619-A 1 01-FEB-2001;  
CONNAUGHT LABORATORIES LIMITED (CA)  
FEATURES  
source Location/Qualifiers  
/organism="Moraxella catarrhalis"  
/db\_xref="taxon:480"  
BASE COUNT 2365 a 1555 c 1533 g 1620 t  
ORIGIN  
Query Match 66.4%; Score 4090.6; DB 10; Length 6973;  
Best Local Similarity 81.3%; Pred. No. 0;  
Matches 5063; Conservative 0; Mismatches 1009; Indels 157; Gaps 21;  
QY 1 atgaatcacatctataaagtcattcttttaacaaagccacagggcacatttatg99cgtg99g 60  
Db 542 ATGAATCACATCTATAAGTCACTTTTAAACAAGCCACAGGCACATTTATGCGCATTGCA 601  
QY 61 gaatagcgaatccccacagtac-gggggggggtagctgtgtacagggcgaagtgtgacag 119  
Db 602 GAGTACGCCAAATCCACAGCAGCGGGGGGTAGCTGTGCTACAGGGCAAGTTGGCAG 661  
QY 120 tgcagcaacttaagctttgcccgtattgcccgcctcgtctgctcgtgctcgtgctcgtc 179  
Db 662 TGATATGACTCTGAGCTTTGCCCGTATTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 721  
QY 180 gctcaatggcagtgctttagctcaacaaattactacca---agatcgaatattgttcaaac 236  
Db 722 GCTCAGTGGCAGTGCTTATGCTCAAAAAAAGATACCAACATATCGCATTTGTTGAACA 781  
QY 237 aaacaagataaacaacagcgtgaaagcgtgcccgaagcgtgaaagcgtgaaagcgtgaa 296  
Db 782 AAACAGGCCAACAGCGCTC-----AGGCAGTCCCAAGCGGCGGTGATCGAGCCATTGC 835  
QY 297 ttltggtagcttttctaaggcacaaggctctcaagctatttgcctatcgttagtgtagtaaac 356  
Db 836 TATTGCTGAATAATGCTAACGCACAGGGGGGTCAAGCCATCGCCATCGGTAGTAGTAATAA 895  
QY 357 agatccctaataatgtagtaataatgtaggtttccacgcacaaaggtaaacgagtcctat 416  
Db 896 AACTGTCAATGGAAGCAGTTTGGATAAGATAGGTACCGATGCTACGGGTCAAGAGTCCAT 955  
QY 417 cgcctcgtggtgtagtattggtgaggggtgagctgctcgtgagtcgcatcggtgagtgatga 476  
Db 956 CGCCATCGGTGGTGTGTAAGGCTAGTGTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 1015  
QY 477 ctlatattgcttaagaatcttgatctgaagaa-----tgaattttccaa 521  
Db 1016 CTTTACATTTGCTTGTGATCAGCATGGTAACTCTAAACATCCGAAAGGTACTCTGATTAA 1075  
QY 522 acttattcatggccatgaataattaaaaaaalacaaacctcaacccgagtggttaaatcaa 581  
Db 1076 TCTATTAAACGCCATCGACGTATTAAAGAAATACGAAGCTCAAGAGATGAATGTGTA 1135  
QY 582 atatcgacgacaagagcacaagggcagccagctactgcaagtgggagccatgctcatatgc 641  
Db 1136 ATATAGACGCACACCGCAAGCGGACACGCCAGTACTGCAAGTGGAGCCATGTCATATGC 1195  
QY 642 acagggtcattttttcccaacgcctttggtatcacgacaacagctgaagctgcctattcctt 701  
Db 1196 ACAGGGTCATTTTCCAAACGCCCTTTGGTACACGGGCAACAGCTAAAGTGCCTATTCC 1255  
QY 702 ggcagtaggtcttgcgcgcgcacaaagccacaaatctcaatcgctgttgggttccaatgc 761



Db	1256	GGCAGTGGGTCTTTCGCCGCACACCGCAGGGCCAAATCTACAAATGCTATTTGGTTCTTGATGC	1311
Qy	762	aaaagctaaacgcgttttcagcagcacagccatttggtggaataactctagtttaatttgggtcg	821
Db	1316	AACATCTAGCTCGTTGGGAGCGATAGCCCTTGGTGCAGGTACTCGTCTCAGCTACAGGG	1375
Qy	822	aggogttgccttaggttttggttctcagataccttgataggtggataataatacagatgccag	881
Db	1376	CAGTATTGCCCTAGGTCAAGGTTCTGTGTCTACTCAGAGTGATATAATATTC---TAGACC	1432
Qy	882	tgccatgtaccactaggtaaaacgttagcagacagcatataaagccaccgcgcagggtga	941
Db	1433	GGCCTATACCAAAATPCCCGACCACTAGACCCCAAGTTTCAAGCCAC---CAATAATAC	1489
Qy	942	tctcagcatatatttccatctggaataagtaataataataatagcagtatcaggcgtaa	1001
Db	1490	GAAGCGGGTCCATTTCCTATTGGTAGTAA-----CTCTATCAACCGTAA	1534
Qy	1002	aatcatcaatgctgcgttcggtttcttcgggatcacgatcggtcctaattgtgcacagcttaa	1061
Db	1535	AATCATCAATGTCGGTGCAGGTGTTAATAAAAAACCGATCGCTCAATGTGGCACAGCTAGA	1594
Qy	1062	attggtggaggaactggcta---atcgtaaaattacttttaagggatgagtgacaataa	1118
Db	1595	AGCGGTGTCGAAGTGGGCTAAGGACGCTAGAAATTACTTTTCAGGGTGATG-----ATAA	1648
Qy	1119	tagcaatagcgtagaagaggttttggtcaatactttaactattaaagtgatgcacagc	1178
Db	1649	CAGTACTAGCTAAAAATAGGTTTGATTAATTACTTTTAATTAATTAAGGTGGTGCAGAGAC	1708
Qy	1179	caacgcattaacccaagctaacatcggtgtggtgaacagatggcaa-----tggtctgaa	1232
Db	1709	CAACGCATTAACCGATAATAATATCGGTGTGGTGAAGACAGGCTGATATAATAGTGTCTGAA	1768
Qy	1233	agttaaacttgctaaagagctgactggattgaccagtgtctcc-----gc	1277
Db	1769	AGTTAAACTTTGCTTAAACCTTTTAAACAATCTTACTGAGGTGAATACAACTACATTTAAATGC	1828
Qy	1278	taccacaataatcacggttagtaataaccacaacaacgccagctacaaagcgttg	1337
Db	1829	CACAAACCACAGTTAAGGTAGGTAGTAGTAGTAGTACTACAGCTGAAATTTAGTGTATAG	1888
Qy	1338	tttgaccttagccca-----atacaggtgtacaaaaacagataaaacggtctacagcat	1391
Db	1889	TTTAAACCTTTACCCAGCCCAATACAGGCAGTCAAAAGCACAAAGCAAAACCGTCTATGGCGT	1948
Qy	1392	tgatgattgaadttaactaaatgataagtaataagtagtaacgaactaaagttactactgatat	1451
Db	1949	TAAATGGGGTGAAGCTTTTACTAAATTAATGCAGAAACACACAGCAGCAATACGCGAC'TACTCGTAT	2008
Qy	1452	taccaaaaagaaaattggttttgctggtaactaatgatggagtgtatgaagaagcaaacctta	1511
Db	2009	TACCAGAGATAAAATTTGGCTTTGCTCGAGTGGTGATGTTTATGAAAAMAAACAGCACCATTA	2068
Qy	1512	tcttgacaacgaagaagctaaaagttggcaacgacccctaaccagtggttagcttgactgt	1571
Db	2069	TTTGGATAAAAAACAATTAAGTGGTAGTGTGCAATTAACCATAGACAATGSCATTGA	2128
Qy	1572	taataaacaccacttggtaataaacaaatccaagtcgggtcctaattggaatttgcac	1631
Db	2129	TGCAGGTAATTAATAAGATCATAGTAATCTTGCCAAAAGGTAGCAGTGTCAACGATGCGGTATAC	2188
Qy	1632	agtcgctaaalaatttgcaaatacctcagcaacagtcggcactgctcgtattaccgaaga	1691
Db	2189	CATCGAACACGCTAAAGCGGCCNAGCCTACTTTTAAACGCGAGGCGTGGCATCTAGTGTAC	2248
Qy	1692	gaaaattggttttgcgttactaatgatggagtgtgaacaaagcaccataatttggataa	1751
Db	2249	ACCTACTGAAATATCAGTTGATGCTTAAGAGTGGCAATGTTACCGCCCCCAACTTACACA-	2307
Qy	1752	agaacgacttaagtggtggctgttgtaaatatcaccacagatagtggttattaatgctgtdaa	1811
Db	2308	-----TTTGGGCTGAAAACACCACCGAGCTTAACAGTGTAGTGCAC'TAGTGATAAATTTAG	2359

QY 2883 cgcctcacaacatttaccgttataaaagtgatgaaatgataaggctgtagacacaa 2942  
|||||  
Db 3398 CGCCCTACAAACCTTTACCGTTTAAAGAGTAGTGAATAATAATATGCTGATGACGCCAA 3457  
QY 2943 cgcctacacccgtgggttaagaatggaacaaagtggttaaaatcaacaccttaaaactcaaaagg 3002  
Db 3458 CGGCATCACCGTGGGTCAAAAGAACGCAAAATCAATCAAGTCAACACCCCTAACACTCAAAAGG 3517  
QY 3003 taaaaacgctcttgatattaaacagacaaagatggttacggtttacctttggcattaaacac 3062  
Db 3518 TGAACAGGGCTTAAATATTAAACCCGACAAAATAATGTTACCTTTGGCATTAACAC 3577  
QY 3063 ccaaagcggctttaaagccggcgagacagcaccactctaaacaacaatggtctgtctattaa 3122  
Db 3578 CACAAAGCGGTCTTAAAGCGGGCAAAAAGCACCC--CTAAACGACGGTGGCTTCTGCTATTAA 3634  
QY 3123 aaacacccgttagtaacgaacaaatcaaatcggtgctgtagtgggaagtttggccatgg 3182  
Db 3635 AAACCCACCTGGTAGGCAACAAATCCAAAGTGGTGTGATGGCGTGAAAGTTGCCAAGGT 3694  
QY 3183 ---taataatgggttaggtgctggcattgtaggacaaactcgaatcgaatcgaatcgaatcga 3239  
Db 3695 TAATAATATGGTGTGTAGGTGCTGGCATTTGATGACAACTTCGCATTTACAGAGATGA 3754  
QY 3240 aattggcttactgggactaatggctcacttgatgataaagcaaacccacccactgaagcaaga 3299  
Db 3755 AATTGGCTTTACTGGGACTTAATGGCTCACTTTGATTAAGCAAAACCCACCTTAAGCAAGA 3814  
QY 3300 cggcattaaacagcaggtgtaaaagattaccacattcgaatcgaatcgaatcgaatcgaatcga 3359  
Db 3815 CGGCATTAAACGACGCTGTAATAAGATTTACCAATTTCAATCAAGTCAAGTCAAGTCAAGTCA 3874  
QY 3360 cagccatgactgctgacagcggcgaagatttatgatttaaaacagaaacttgaataa 3419  
Db 3875 CAGGCATGATGCTGTGACAGCGGCAAGATTTATGATTTAAACCGCACTTGAACAAACAA 3934  
QY 3420 aatcagcagctgctcgaacagacacaaactcattacacgaattctcagtagcagatga 3479  
Db 3935 ANTACAGAGTACTGCCAAACAGCACAAACTATTACAGAACTTCTCAGTAGCAGATGA 3994  
QY 3480 acaaggtaataacttaccggttagtaaccttactcagttatgaacactcaagacctc 3539  
Db 3995 ACAAGGTAAATACCTTTACGGTTACTAACCCCTTACTCAGTTATGACACCTTCAAAAGACCTC 4054  
QY 3540 taatctacacacttgcaggtgaaacggcattaccacacgaagtaataaagggtggt 3599  
Db 4055 TGATGTCATCACCTTTGCAGGTGAAACGGCATTTACCACCAAGGTAAATAAAGGTGTGGT 4114  
QY 3600 gcgctggtggcattgacaaacaaaggcttaaccacgcctaaagctgaagctggtggt 3659  
Db 4115 GCGTGTGGGCATTGACCAACCAAGGCTTAACCAAGCTTAAGCTGACCGTGGGTAAATAA 4174  
QY 3660 taatggcaagcattgctattacagcgaataatggtcaaaataccatcaacaggaactaa 3719  
Db 4175 TAATGGCAAGGCATTGTCAATTACAGCCAAATGCTCAAAATACCAATACCAAGGTAAAG 4234  
QY 3720 caacactctagcttaatttaccacatgaataagggtagcgtacgcacacacagacagggcaa 3779  
Db 4235 CRACACTCTAGCTTAATGTTACCAATGATAAAGGTAGCTAGCGACACAGAACAGGGCAA 4294  
QY 3780 tataatcaagacgaagacaaacccgtgcccgaagcattggtgctgaagcgcagg 3839  
Db 4295 TATAATCAAGAGAGAGACAAACCCGTCGCCAGCATTTGATGCTGCTTAAGCGCAGG 4354  
QY 3840 cttaacttgcaggcaatgggtgaagcgggtgacttctccacttatgacacagctcaa 3899  
Db 4355 CTTTAACTTGCAAGGCAATGTTGAGCGGTTGACTTTGCTCCACTTTATGACACCGTCAA 4414  
QY 3900 ctttgccaatggcaataccacacccgtgaagtgacacttatgatgacaaagcaaacag 3959  
Db 4415 CTTTGGCGATGGCAATGGCCACACCGCTAAGGTGACCTATGATGACCAAGCAAAACCCAG 4474  
QY 3960 taaagtgtctatgattcaatgtgtagatatacaacacattgaagttaagataaaaaact 4019

Db 4475 TAAAGTGGCTATGATGTCAATGTGATGATCAACCATTTGAAGTTAAAGATAAAAAACT 4534  
QY 4020 tggcgtataaaacacacacattgacagtagtgacaggtgctlaataataatttgccttaag 4079  
Db 4535 TGGCGTAAAAAACACACCATTTGACCAGTACTGGCACAGGTGCTTAATAATTTGCCCTAAG 4594  
QY 4080 caatcaagctcactggcgatgctgcttcaaggccagtgatataogttgctcatctaaacac 4139  
Db 4595 CAATCAAGCTTACTGGCGATGCGCTTGTCAAGGCGAGTATATGTTGCTCATCTAANAC 4654  
QY 4140 ctatctggcgacatccaaactccaaaggggcaagccagcaacaaactcagcagggcta 4199  
Db 4655 CTTATCTTGGGGACATCCAAACTGCCAAAGGGCAAGCCCAAGACCAAACTGCTGCGCCA 4714  
QY 4200 tgtgtagctgtagtggaataaggtcactatgacagtagccgataaacaagtagctatcaagc 4259  
Db 4715 TGTGGATGCTGATGGCAATTAAGGTCTCTATGACAGTACCGATAACAAGTACTATCAAGC 4774  
QY 4260 caaaaatgtaggcacagcttgataaaacaaaggaagttgccaagacaaactggtcgccca 4319  
Db 4775 CAAAAATGATGGCACAGTTGATAAAACCAAGAAAGTTGGCCAAAGACAAACTGCTGCGCCA 4834  
QY 4320 agccaaacccagatggcagctcaaatgaatgtcaaaatcagctcattcaacaaaga 4379  
Db 4835 AGCCCAAAACCCAGATGGCAGCTTTGGCTCAATGAATGAATCAATCACTCATTAACAAAGA 4894  
QY 4380 acaagtaaatgatgccataaaagcaaggcatcaatgaagacaaaccccttctgtaagg 4439  
Db 4895 ACACTAAATGATGCCAATAAAAGCAAGGCATCAATGAAGACAAACGCTTTGTTAAAGG 4954  
QY 4440 acttgaaagacgcttcttgatacaaaacaaagccagcagtagtaactgtgggtgattt 4499  
Db 4955 ACTTGAAAGACGCTTCTGATGAACAAACCAACCAAGCCGCGAGTACTGTGGGTGATTT 5014  
QY 4500 aaatcgcgtgcccacacacgcgtgaccttgcaggggatacagggacaaacggcttaaaaa 4559  
Db 5015 AANTCGCTTGCCCAACACACGCTGACCTTTGAGGGGATACAGGCACAAACGGCTTAAAA 5074  
QY 4560 actggcgagactttgacatacaaaaggtgggcaaacagacacacataagctaaacgataa 4619  
Db 5075 ACTGGCGAGACTTTGACCATCAAAAGGTGGGCAAAACACACACCAATTAAGCTAACCGATAA 5134  
QY 4620 taactcgcgttggtagcaggtactgtagtcttactgtcaaaacttgcgaagacactaac 4679  
Db 5135 TRACATCGGTGTGTAGCAGGTACTGATGGCTTCACTGTCAAACTTGGCAAGACCTTAAC 5194  
QY 4680 caatctaacagcgttaattgagggtagcagggtagcagggtagcagggtagcagggtagcag 4739  
Db 5195 CAATCTTAACAGCGTTAATGTCAGGTGGCACCAAAATTTGATGACAAAGGCGCTGCTTTGT 5254  
QY 4740 aagcgaacgggtcaagcgaacgaacacacccctgtgctaaagtgcgaatgggctgaacct 4799  
Db 5255 AGACTCAAGCGGTCAAGCCAAAGCAAAACACCCCTGTGCTAAGTGGCAATGGGCTGGACCT 5314  
QY 4800 ggttggaaggttcagtaattggtgggcaaaagggcaaaagatacagacgctgccaagt 4859  
Db 5315 GGCTGGCAAGGTCAATGTTGGGCAAGGCAACAAAGATACCGACGCTGCGCCTTGT 5374  
QY 4860 acaacagttaaacgaagtagcaactgttgggtct-----tggttaatgataacgc 4910  
Db 5375 ACACAGTTAAACGAAGTAGCAGCACTTTGTTGGTCTTGGTAAATGCTGCTAATGATAACGC 5434  
QY 4911 tgacggcaatcaggttaacatctgccacatcaaaaaagaccccaaatcagggttcacatc 4970  
Db 5435 TGACGGCAATCAGTAAACATTTGCCGACATCAAAAAAGACCCAAATTCAGGTTTCATCATC 5494  
QY 4971 taacccgactgtcaataaagcagcagcttgcgggttaaagggttaaagcagataccga 5030  
Db 5495 TAACCGCACTGTATCAAAAGCAGCAGCGTACTTTGGCGGTAAAGGTAAATAACCATACCGA 5554  
QY 5031 aaacttgccactggtggtgtacaagtggcggtggtgataaagacggcgaacgctaacggcga 5090  
|||||



QY 1716 tgatgagtgatgaacagcaccatatttggataaaagaagacttaaaagtggtcggtg 1775  
Db 44112 TGTACAGTTGATGACAAATTCACCATATTTGGATAACGCAAAAGCTTGAAGTTGGCAATGT 44171  
QY 1776 tgaatatccacaaga tagtgggtatttaattgctggttaatacaagattaccggaactactaa 1835  
Db 44172 TGAGATTAATAAAGACACTGGTATTAATGCAGGTAAATCAAAAGATTACCGGACTTACTGA 44231  
QY 1836 tggatatagcaaaataccgatgcgtttaccatcaaacagctcaaaagcgcgaagcctacttt 1895  
Db 44232 TGGTGCAGCAGTGCCTGCTTACCATCAAAAGCTCAAAAGCGCCAGCCTAATTT 44291  
QY 1896 aaacgagcgatggcatcagtttaataatgaataacggggggtacttagtga---tagtag 1952  
Db 44292 AACCCAGGCAATGGCAATTGAGATTAAACATAACAAACAGTCTATTAGTTGATGCTGCTAA 44351  
QY 1953 tggcaatatatacccccactataacattagctgtaaaacacactaaagcttaacagtaa 2012  
Db 44352 TGGTAATGTTACCCGCCCTCTTATACCATTTGGGTGAAACCCACCACCTCTGACCAATAA 44411  
QY 2013 tggcaccagtggttaataataatttagtggtaattgctcatgataaacaataagcttagt 2072  
Db 44412 TGGCCCTAGTGTAGTAAATTTGCTTTAAGGATAGTAAATACGCATAAACAAATAGCTTAGT 44471  
QY 2073 taccgcaaaagatttggcagactatataaataaagtcacaaagcgcgtgacagtgctct 2132  
Db 44472 TACGCCCAAAAGATTGGCAGGCTATCTAATATGAAGTCAATCGAACGGCTGACAGTGCTCT 44531  
QY 2133 accaagctttaaaglcacaaacgggtgataatagcaaacacgcaccccggtgggtaaaaga 2192  
Db 44532 ACAAGCTTTACCGTTAAAGGTGGTGTGATGCAGCAAGCAACACAGCATC----- 44580  
QY 2193 tacaacggcgaagacttcaacaccttaaaactcaaaaggtgaacgcgtgttaatatattac 2252  
Db 44581 -ACGCTAGATAAAACACGACAGTCTAAACGCTAGCAGCGGTAAACGGCATTACTGTATAA 44639  
QY 2253 gaccataagagccacaggtacagttacaccttggcattgaccacaaagttatggtctcacac 2312  
Db 44640 AACGATACACTCAGAAAAAAGTCACGGTTGGCATTTGACCAAGCCAAATGCTTAAGCAC 44699  
QY 2313 gacctagctgacggtgggttagcgatacaaaatggtaacgaltggttattgagcaagtcct 2372  
Db 44700 GCGTAAGCTTGACCGTTGGGTAGCAATAGTAA---AACTCAATTTGGTTATTGACCAA---- 44751  
QY 2373 tagcgtgacggttaacagcaccaaaaacatcatataaaggattgtccccaacactgcctag 2432  
Db 44752 -ATGTTATCGGTACGACACCAAAACATCATCAAAAGGCTGTCTCTCAACACTGACTGA 44810  
QY 2433 cattgcagtcgaagtgccgcgaacatagcactggcgcaatacaatcgaaagaaaagacaa 2492  
Db 44811 CATTACCAACACAGATTAATACACACACACCAACAGCAATGACGCCCAAGGCANAGA 44870  
QY 2493 atccaacgctgccagcatgagtgtctaaatgcaggcttttaaccttaaaaaataatggtg 2552  
Db 44871 ATCCAACGCTGCCAGTATCAAGATGCTGCTAAACGCAAGCTTTAACTTCAAGGCAATGG 44930  
QY 2553 caaagacaaagacttggctccacttatgacactgttgactttatcgatggcaatgcccac 2612  
Db 44931 CAAGCGGTTGACTTGTCTCTACTTATGATACCGTCAACTTTGGCAATGGCGATGGCAC 44990  
QY 2613 caccgcacacagtaac-----ttatgatgaagccaatcaaacagtaaaagtggtgctatga 2666  
Db 44991 CACTGCCACAGTACCTTATANTTATATGANGGCAAAACCAACCAACGACGACGATGCTATAA 45050  
QY 2667 tgtgaattggatgagaaaaacattgaaactgacagcgagataatggcaagaataaacttgg 2726  
Db 45051 TGTCAATGTGCATGATACAACCATTCATCTGGCAAAATAATATGGTAAATAAAACCTTAG 45110  
QY 2727 cgtcaaaacacatcaaacctgacgaacaa---gtactaatggtaatgcaactacatttag 2783  
Db 45111 CGTAAAAACCAACCATTTGACCAAAACAAATATGTTGCAAAATGTTGAATGAATCAACTAAATTTAG 45170  
QY 2784 taccgacgatgacctgaccttggtaagcgaagtgtatcgccggcgaataataaacacct 2843



QY	4992	agggcaggtacttgccggtataaaggaataacgatacagcaaaacttgccactggtggtg	5051
DB	47364	AGGCACGGTACTTGGCGGTAAAGGTAATACGATACCGAANAATTCGCCACTGGTGTAT	47423
QY	5052	acaagtggcggtggtataaagcagcgaacgctaagcgcgatttaagcgaatgtttgggtcaa	5111
DB	47424	ACAAAGTGGCGGTGGATAAAGACGGAACGCATAACGGCGATTTAAAGCAATGTTGGGTCAA	47483
QY	5112	aacccaaaagatggcagcaaaaagccctgctccacttataacgcgcaggtcagac	5171
DB	47484	AACCCAAAAGATGGCAGCAAAAAGCCCTGCTGCCACTTATACCGCGCAGGTCAGAC	47543
QY	5172	caactatgaccaaaccgcaagcgaagcattgacagaataaatagaacaaggtatccg	5231
DB	47544	CNACTATTGACCAACACCCCGCAGAGCCATTGACAGAAATAATGAACAGGTATCCG	47603
QY	5232	ctttctccatgtcaacgatggcaatacaagacctgtgtgtacagggcgttaacggcatgga	5291
DB	47604	CTTCTTCCATGTCACGATGGCAATCAAGAGCGTGTGGTACAGGCGGTAAACGGCATTTGA	47663
QY	5292	ctcaagtccctcaggaagcactcagtgccgaaggtttccagggccaaagcgagatgggtga	5351
DB	47664	TTCAAGTGTCTGCTGGCAAGCAATTCAGTGGCGATAGTGTCCAGGCCAAGCGAGATGTCGA	47723
QY	5352	agccgctgttgccatagcagcaaaaccccaagcagcgaacccaatccatccatcgtgtga	5411
DB	47724	AGCCGCGTGTGCCATAGGCGAGACAAACCCCAAGCAGCAGCAATCCATCGCCATCGGTGA	47783
QY	5412	taagcgaacacccagcgatcaatccatccgcaacggtacagggcaatgtgtgtagcagg	5471
DB	47784	TAAAGCAACGACCGCGGCGATCAATCCATCGCCATCGGTACAGGCAATGTGTGTAGCAGG	47843
QY	5472	taagcactctgtgcccacggcgacccaacacactgttaaggtgataaacttaccagtgt	5531
DB	47844	TAAACACTCTGTGTGCCATCGCGGACCCCAAGCACTGTTAAAGGCTGATTAACAGTTTACAGTGT	47903
QY	5532	gggtataaacaacagtttacgctgacactcaaacgcatgtctttggtgggcaataa	5591
DB	47904	GGGTAAATACCAACCACTTTACCGATGCCATCGCACTACAGCCGATGTCTTTGGTGGCAATAA	47963
QY	5592	catcacgtgacggaagtaactcgtgtgctttaggttcaaacctcctccatcgtcagg	5651
DB	47964	CATCACGCTGACCGAAGTAATCTCGGTGGCTTAGGTTCAGTTCCTGCTGCTGCTCCGT	48023
QY	5652	cacacacgagcagcacagcaaaactcgaacgacagcagcaggtacacacacacagc	5711
DB	48024	CACACACGAGCAGCACAAAGCCCAAAAATCTGACGGCAGCAGCAGGTACAAACACACAGC	48083
QY	5712	aggtgcccaggtacggttaaggtttgctggcaaacgcggttgggtcgggtcctcgt	5771
DB	48084	AGGTGCCACAGGTACGGTTAAAGGCTTTGCTGGACAAACGGCGGTGGTGGGCTCTCCGT	48143
QY	5772	gggtgctcaggtgtcgaacgctgtatccaaatgtggcagcaggtgaggtcagtgccac	5831
DB	48144	GGGTGCTCAGGTGTGACCGCGGTATCCAAAATGTGCCACAGCAGTGAGGTCAAGTCCAC	48203
QY	5832	cagcacgatgctggttcaatggttagcagttgtacaaagccaccccaagcattgcaacgc	5891
DB	48204	CAGCACGATGCGGTCAATGGTAGCCAGTTGTACAAAGCCACCCCAAGCAATGGCCACGC	48263
QY	5892	aaccaaagacttgacatcgtatccacaaacgaataaagcaatcgagggatttc	5951
DB	48264	AACCAATGAGCTTGACCATGATTCACCAACCAACCAACCAATTAAGCAATGACGAGGATTC	48323
QY	5952	atacagatggcgatggcggtccatgccaagcctacattcctgagatccatggttac	6011
DB	48324	ATCAGCGATGGCGATGGCGTCCATGGCCAAAGCCTACATTCCTTGGCAGATCCATGGTTAC	48383
QY	6012	cgggggtattgccaacacacagcgttaaggtcgggtggcgatgggactgcgaagctgac	6071
DB	48384	CGGGGGTATTGCCACACCCACCAACGGTGAAGGTGGGTGGCGAGTGTGCGAGCTGTCGAAGCTGTC	48443

QY	6072	ggataatggtcaatgggtatttaaaatcaatggttcagccgatacccaagccatgtagg	6131
DB	48444	GGATATATGTCATGGGTATTTAAATCAATGGTTACGCCGATACCCAAAGCCATGTAGG	48503
QY	6132	ggcggcagttggtgcaggttttcaacttt	6159
DB	48504	GGCAGCAGTTGGTGCAGGTTTTCACCTTT	48531
RESULT	7		
AX079914	AX079914	5976 bp	DNA
LOCUS	Sequence 2 from Patent WO0107619.		PAT
DEFINITION	AX079914		
ACCESSION	AX079914		
VERSION	AX079914.1	GI:13159437	
KEYWORDS	Moraxella catarrhalis.		
SOURCE	Moraxella catarrhalis.		
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;		
REFERENCE	1 (bases 1 to 5976)		
AUTHORS	Loosmore, S.M., Sasaki, K., Yang, Y.P. and Klein, M.H.		
TITLE	Recombinant high molecular weight major outer membrane protein of moraxella		
JOURNAL	Patent: WO 0107619-A 2 01-FEB-2001;		
FEATURES	CONNAUGHT LABORATORIES LIMITED (CA)		
source	Location/Qualifiers		
	1. 5976		
	/organism="Moraxella catarrhalis"		
	/db_xref="taxon:480"		
BASE COUNT	1972 a 1360 c 1355 g 1289 t		
ORIGIN			
Query Match	64.1%;	Score 3947.8;	DB 10; Length 5976;
Best Local Similarity	80.9%;	Pred. No. 0;	
Matches	4905;	Conservative 0;	Mismatches 1002; Indels 156; Gaps 20;
QY	166	gtgacgtgctgagcgtcaatggcagtgcttctgctcaacaaattactacaag---atc	222
DB	1	GTGATCGGTGCNCGCTCAGTGGCAGTCTTATGCTCAAAAAGATACCAACATATC	60
QY	223	gaattgtgtcaacaaacagataaaacacgcgtgaagcgatgccctagcgacaggt	282
DB	61	GCAATGGTGAACAAACACGACCAAGACGCTC-----AGGCACTGCCAGCGGCGGT	114
QY	283	gaagcatcctgcttctggttagtcttcttaaggcacaaggtctctcaagctatgctatc	342
DB	115	GATCGAGCCCTTGCTATGTTGGTCAAAATGCTAACGCACAGGCGGTCAAGCCATCGCCATC	174
QY	343	ggtagtgtaacacacagatcctaataatggttagtaatggttaggttccacgccaac	402
DB	175	GGTAGTAGTAATAAAACTGCTCAATGGAAGCAGCTTTGGATAAGTAGGTACCGATGCTACG	234
QY	403	ggtacagagtcctacccatcgctggtggtggtggtggtggtggtggtggtggtggtggtg	462
DB	235	GGTCAAGAGTCCATCCCATCGGTGCTGATGTAAAGGCTAGTGTGATGCTGCTGATGCTG	294
QY	463	atcggtagtgatgacttatatttgcctcaagaatcttgatctgaaagaa-----	509
DB	295	ATCGGTAGTCATGACTTACATTTGCTTGTATGATGATGATGATGATGATGATGATGATG	354
QY	510	--tgaatttcaacaaacttattcgtggtggtggtggtggtggtggtggtggtggtggtg	567
DB	355	ACTCTGATTAACGATCTTATTACGGCCATGCGATATTAAAGAAATACGAAGTCAAAAG	414
QY	568	gattgtaaaatcaaatatcagcagcaacagcagcagcagcagcagcagcagcagcagcag	627
DB	415	GATAATGATGTAAATATAGACGCACAAACGCGACACGCGGACACGCCAGTACTTGCGTGGGA	474
QY	628	gcaatgcatatgcaagaggtcatttttcaagccttttgcgtacacacacacacacacacac	687
DB	475	GCCATGTATATGACAGGGTTCATTTTCCACGCCCTTTTCCACGCCCTTTTCCACGCCCT	534







Thu Sep 13 14:19:03 2001

Db	2617	ACGCCAAAGACATCGCCGAAATCTAACACCCCTAGCCAAAGGAAATTCACACCACCAAA	2676	3694	TATGACACCGTCAACTTTGCCGATGGCAATGCCACCACCGCTAAGGTGACCTATGATGAC	3753
Qy	2869	ggcagcaaacacacgcgccttacaaacctttaccgttaaaaggtagatgaatgaatgaag	2928	3946	acaagcaaacacagtaaaagtgtctctatgatgtaagtgtgatatatacaaacattgaagtt	4005
Db	2677	GGCAGCAGCAGACACGCCCTTACAAACCTTTACCGTTAAAGAGGTAGATGAATAAT	2736	3754	ACAAGCAAAACCGATTAAGTGTCTATGATGTCAATGTGGATGATACAAACCTGAAGTT	3813
Qy	2929	gctgatgacacaaacgcacacccgtggtgaagatggcaacagtggttaagtgcaaac	2988	4006	aaagataaaataaacttggtggttaaaacccaccacattgaccagttactggcacaggtgctaat	4065
Db	2737	GCTGATGAGCCANCCGCAATCACCGTGGTCAAAAGACGCAAAATTAATCAAGTCAACACC	2796	3814	AAAGATAAAATAAATTGGCGTAAACACACACATTTGACAGTACTGGCACAGGTGCTTAAT	3873
Qy	2989	ttaaactcaaaagttaaaacccgtcttgatataaaacgcacaaagatggtacggttacc	3048	4066	aaatttgccctaagaacaataaagctactggtggtggtggtggtggtggtggtggtggtggt	4125
Db	2797	CTAACACTCAAGGTGAAACCGCTTAATAATTAACACCGCAAAATTTAGTACGTTACC	2856	3874	AAATTTGCCCTAAGCAATCAAGCTACTGGCGATGGCTTGTCAAGGCCAGTGTATCGTT	3933
Qy	3049	tttggaattaaacccaacgaagcgtctaaagccgcgcgcacacacacacacacacac	3108	4126	gctcaactaaacaccccttatctggcgacataccaaactgccccaaaggggggcaagcaaac	4185
Db	2857	TTTGGCATTAACACCAACGCGGTCTTAAGCGCGGCAAAAGCACC--CTAAACGACGCT	2913	3934	GCTCATCTAACACCTTATCTGGCGACATCCAACTTGCNAAGGGGCAAGCCAAAGGAC	3993
Qy	3109	ggcttgctataaaacacacccgttagtaacgaacaaataccaaagtcggtgctgacggtg	3168	4186	aactcagcaggtctatggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt	4245
Db	2914	GGCTGTCTATTAACACCCACTGGTGTAGCAAAATCCAAAGTGTGCTGATGGCGTG	2973	3994	AACTCAGCAGGCTATGTGGATGCTGATGGCNATTAAGTCTATCTATGACAGTACCGATAAC	4053
Qy	3169	aagtttgcactggt---taataatggtttagtgctggtggtggtggtggtggtggtggtggt	3225	4246	aagtactatcaagccaaataatggtggtggtggtggtggtggtggtggtggtggtggtggt	4305
Db	2974	AAGTTGCCAAGGTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3033	4054	AAGTACTATCAAGCCAAATGATGGCACAGTTCATATAAAACCAAGAAAGTTGCCAAGAC	4113
Qy	3226	attaccagagatgaattggttactgggactaatggtcactggtcactggtcactggtcactggt	3285	4306	aaactggtgcgcgaagccaaac	4365
Db	3034	ATTACCAGAGTAAATTTGGCTTTACTGGGAC"AAATGGCT"CACTTGTATAAAGCAAAACC	3093	4114	AAACTGGTGCGCCAAGCCAAACCCAGATGGCNATTAAGTCTCAATGATGTCAATCA	4173
Qy	3286	caactaagcaaacgacattaaacgaggtggttaaaagattaccacacattcaatcaggt	3345	4366	gtcattcaaaagaaacaaagtaaatggtggtggtggtggtggtggtggtggtggtggtggt	4425
Db	3094	CACCTAAGCAAGAGCGGATTAACGAGGTGGTAAAGATTAACCAATTTCAATCAGGT	3153	4174	GTCATTACAAAGAACAAAGTAAATGATGGCAATTAAGAACAGGCATCAATGAAGACAC	4233
Qy	3346	ggatgtgcacaaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	3405	4426	gaccttgtaaaagacttgaaaaagccgcttctgatacaaaacacacacacacacacacac	4485
Db	3154	GAGATTGCCAAACACAGCATGCTGTGACAGCGCGCAAGATTTATGATTTAAACAC	3213	4234	GCTTTGTTAAAGGACTTTGAAAAAGCCGCTTCTGATTAACAAAAACCAAAACGCCGACGA	4293
Qy	3406	gaactgaaaaataaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	3465	4486	actgtgggtgatttaaaatggtggtggtggtggtggtggtggtggtggtggtggtggtggt	4545
Db	3214	GAACTTGAAACAAATCAGCAGTACTGCCAAACACAGCACAACCACTCATTAACGAATTC	3273	4294	ACTGTGGGTGATTAATGTCGCTTCCCAACACCGCTTGACGAGGATACAGGC	4353
Qy	3466	tcagtacagatgaacaaagtgataaactttacgttggttagtaaccccttactcagttatgac	3525	4546	acaagcgtcaaaaactgggagagacttgaccactcaaggtgggcaaacagacacacac	4605
Db	3274	TCAGTAGAGAGTGAACAGGTAAATTAATTTAGGTTAGTAAACCTTACTCAGTTATGAC	3333	4354	ACAACGGCTTAAAAACTTGGCGAGACTTTGACCATCAAAAGTGGGCNAACAGACCAAT	4413
Qy	3526	acctcaagacacctgattgac	3585	4606	aagctaaccgataataac	4665
Db	3334	ACCTCAAGACCTCTGATGTATCATACCTTTGCGGTGAGGTGAAACGGGATTTACCAAGGTA	3393	4414	AAGCTAACCGATTAATTAACATCGTGTGTGTGAGCAGGTACTGTGCTTCAAACTT	4473
Qy	3586	aataaagggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt	3645	4666	gcaaaagacctaaccac	4725
Db	3394	ANTAAAGTGTGGTGGTGTGGCATTTGACCAACCAAGGCTTAACACGCTTAAGCTG	3453	4474	GCCAAAGACCTAAACCAATCTTAACAGCGTTTATGACGTTGGCCNCCNATTTGATGACAA	4533
Qy	3646	acogtgggttaataatggcaagcattgattacacagcacaataaggttcaaaatacc	3705	4726	ggcatctctttagacgcacaaacggtcaagccaaagcaaacacacacacacacacacacac	4785
Db	3454	ACCGTGGGTAAATAATGCAAGGCAATTTGATTCAGCCCAAAATGGTCAAAATACC	3513	4534	GGCGTGTCTTTGTAGACTCAACGGTCAAGCCAAAGCAACACCCCTGTGTAAAGTGCC	4593
Qy	3706	atcacagactaagcaac	3765	4786	aatggcgtgac	4845
Db	3514	ATCACAGGACTAAGCAACACTAGCTAATGTTTACCAATGATAAAGGTAGCGTACGAC	3573	4594	AATGGCTGGACCTGGGTGCAAGGTCATCAGTAAATGTGGCCNCCNATTTGATGACAA	4653
Qy	3766	acagaacagggttaataataaagacgaagacaaacacacacacacacacacacacacac	3825	4846	gacgtgccaatgtacaac	4896
Db	3574	ACAGAACAGGCAATATAATCAAGAGCAAGCAAAACCCGTCGCCAGCATTTGTTGAT	3633	4654	GACCGTGCATGTACAAAGTTAAACAGTAAACGAAAGTACGCAACTTGTGTGCTTGTGTAATGCT	4713
Qy	3826	gtgctaaagcagcgtttaaacttgaaggtgaaggtggtggtggtggtggtggtggtggtggt	3885	4897	ggtaataac	4956
Db	3634	GTGCTAACGCGCAGGCTTTAACTTCAAGGCAATGGTGAAGCGGTGACTTTGCTCCACT	3693	4714	GGTAATGATAACGCTGCGCAATCAGGTAAACATTGCCACATCAAAAAGACCCCAAT	4773
Qy	3886	tatgac	3945	4957	tcaggttcatcatcaaac	5016
Db				4774	TCAGGTTCATCTATACCGCACTGTATCAACAGCAGCGGTACTTGGCGGTAAAGGT	4833



QY 689 ctgacctattcttgccagtagcttctgcccgaagcacaacaaactcttcaatcgtg 748  
 Db 1589 CAACTTACAAACA-----TTGGCGTGAAACACACCGAGCTTAACAGTGTGCACTA 1639  
 QY 1799 ttaagtctggttaatacacaagattaccggaacttaactaaatggtatagcaataaccgagtcg 1858  
 Db 1640 GTGATAAATTTAGTGTAAAGGGTAGTGTACGAACAATAGCTTACCTCCGCGCAACATTT 1699  
 QY 1859 ttacacataaagcgtccaaagagcccaagcctactttaaacgcaggcgatggcaccagta 1918  
 Db 1700 TGGCAAGCTATCTTAATGAAGTCAATCGAACGCTGACAGTCTCTACAAAGCTTTACCG 1759  
 QY 1919 ttaagttaataacgggagcttagttgtagtagtggcgaatattaccaccccaactata 1978  
 Db 1760 TTAAGAAGAACGATGATGACG-----CCAAAG 1789  
 QY 1979 acattagcgtgaaaaaccacctaagcttaacagtaatggcaccagtggaataataataat 2038  
 Db 1790 CTATCACCGGTGGCTAAAGATACGACAAAANATGCCGGCGAGTCAGCACTTTAAACTCA 1849  
 QY 2039 gtttagtaatgctcatgaatacaatagcttagttaccgcaccaagatttggcagactatc 2098  
 Db 1850 AAGTAAACCGGTCTAACGGTTGCTACCAAAAAGATGGTAGCGTTACCTTTGGGCTTA 1909  
 QY 2099 taaataagtcgaatgaagcgtgacagtgctctcaagctttaagcttcaaaacggtg 2158  
 Db 1910 GCCAAGATAGCGTCTGACCATTTGGCAAAAGACCCCTTAACACAGTGGCTTGACTGTTA 1969  
 QY 2159 ataataagcaaacgcctacacccgtgggtaaaagatacaaaacgcagccctcaaacct 2218  
 Db 1970 AAGTAAACCGAACAACTCAAGTCGGTCTTAATGGCATTTAAATTTACTTAAATGTCGAATG 2029  
 QY 2219 taaactcaacgaagtgaagcgtggttaataattacgaaccaatagagccacaggtacagtt 2278  
 Db 2030 GTAGTAATCCAGGTACTGGCATTTGCNAATACCGCTCGCATTTA---CCAGAGATAAATTTG 2086  
 QY 2279 ctttggcattgacaaagttaagtctcaccacgcttaagctgacccgtgggtagcgata 2338  
 Db 2087 GCTTTGCTGGTCTGTAGTGGTGCAGTTGTATC-----AAACAAACCTTATCTTGATCAA 2139  
 QY 2339 caaatggtaactgattggttattagcaagtccttagcgcgtgacggtaaacagaccacaaa 2398  
 Db 2140 GACAAAGCTA--CAAGTTGGCAATTTAAGATTTACCAACACTGGCATTTAACCCAGCGTGTA 2197  
 QY 2399 acatcattaaagattgcccacacactgcttagcattgcccagtcgaagtgccgcgaaca 2458  
 Db 2198 AAGCCATCAGAGGCTGTCCTCCACACTGCTAGCATTTGCCGATCAAAAGTAGCGCAACA 2257  
 QY 2459 tagcactggcaatacaatacgaagaaaaagcaaatcccaacgctgcccagcattgagtg 2518  
 Db 2258 TAGAACTGGGCAATACAATCCAAAGACAAAGAAATCCCAACGCTGCCAGCATTAATGATA 2317  
 QY 2519 tgcataatgcaagctttaaacttaaaataatggaagacaaagactttgtctccactt 2578  
 Db 2318 TATTAATACAGGCTTTAACTTAAATAAATAATACACCCCAATTTGACTTTGTCTCCACTT 2377  
 QY 2579 atgacactgttgaacttatacgaatggcaatggccacacccgcagtaactatgatgaag 2638  
 Db 2378 ATGACATTGTGACTTTTGCAATGGCAATGGCCACCCAGTAAACCCATGATACCG 2437  
 QY 2639 ccaatcaaacccagtaagtgccgtatgtagtgaatgtagagaaacacattggaactga 2698  
 Db 2438 CTAAACAAAACCGTAGTAAGTGGTATGATGTGANTGTGANTGTGANTACACCATTCATCAA 2497  
 QY 2699 caggcgataatggcaagaaacacttggcgtcaaaacacatacaactgacccgaaacagta 2758  
 Db 2498 CAGGCACTGATGACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2557  
 QY 2759 ctaatggtaatgcaac-----tacatttagtaccagcagatgacctgtgctt 2809  
 Db 2558 CTAAATGTAATACAGCAACTTAATTTAATGTTAATCTTCTAGTGTGANTGATGCTTGTGTA 2617  
 QY 2810 aagccagtgatagtcgcccggcaatctaaacaccccttagccgagggaattcaacaccccaag 2869

QY 689 ctgacctattcttgccagtagcttctgcccgaagcacaacaaactcttcaatcgtg 748  
 Db 536 GTGCCATATTCCTTGGCAGTGGGTCTTGCCGCCACAGCGGAGGCCAATCTACAAATCGCTA 595  
 QY 749 ttgggttccaaatgcaaaagctaacgcgttttgccagcagcagccatgggtggaaataactgtag 808  
 Db 596 TTGGTCTTCATGCAACACTAGCTCGTTTGGGAGCGATAGCCCTTTGGTGCAGGTACTCGTC 655  
 QY 809 ttaatttgggttcgaagcgttgccttagtttggatttccagatccttgatagggataata 868  
 Db 656 CTCACGCTACAGCGCGCTATTCCTTAGGTCAAGGTCTTGTTGTCACACAGACTGATTAATA 715  
 QY 869 atacagatgcagtcgctctatgtaccactaggtgaaacgcttagcagacacagttataagcca 928  
 Db 716 ATTC---TAGACCGGCTATACACCAATATCCAGGCACTAGACCCCAAGTTTCAAGCCA 772  
 QY 929 ccgcgcaggtgattctcaggtatattttccattggttaataataataatagca 988  
 Db 773 C---CAATTAATACGAAGCGCGGTCCACTTTCCATTTGG-----TAGTAAT 814  
 QY 989 gtaacagcgttaaaatcatcaatgctggtggtggttctcgggataccgagtcggttcaatg 1048  
 Db 815 CTATCAAAACGTAAATATCATCTGTCGTCGAGGTGTTAATAAACCAGTACCGTCAATG 874  
 QY 1049 tggcagcagcttaaatggtggggaactggtgta---atcgtaaaattactttaaggtg 1105  
 Db 875 TGGCAGAGCTAGAAAGCGGTGGTGAAGTGGGCTTAAGGAGCGGTAGAATTTACTTTTCAGGGTG 934  
 QY 1106 atggtgacataatagcaatagcgtgagaagaggttttgggcaactctttaactattaag 1165  
 Db 935 ATG-----ATAACAGTACTGACGTAAATAATAGTTTGGGATTAATCTTTAACTATTAAG 988  
 QY 1166 gtatgcacagcaacgcattaaacgaagctaaacatcgcgtggtgtaa-----cagatg 1219  
 Db 989 GTGTCGAGAGACCAACGCAATTAACCGATAAATAATATCGGTGTGGTAAAGAGGCTGATA 1048  
 QY 1220 gcaatgctgaaagttaaaactgtcaaaagagctgctggtgatt-----ga 1264  
 Db 1049 ATAGTGTCTGAAAGTTAAACTTGTCTAAACTTTAAACATCTTAAACATCTTACTGAGGTGAATACAA 1108  
 QY 1265 ccagtgctcgcctaccacaaataacccgtttagtaataaccaacacacacacgcgcagc 1324  
 Db 1109 CTACATTAATATGCCCACACACAGTTAAGGTAGGTAGTAGTAGTACTACAGCTGAAT 1168  
 QY 1325 tacaagcgtggtgtgacacttagcc-----caataacaggtacaaacacagataaaa 1378  
 Db 1169 TATTGAGTGTAGTTTAACTTTTACCAGCCCAATACAGCGCTCAAAAGCACAAGCAAAA 1228  
 QY 1379 ccgtctcagcattgattggaagtttactaataatagtagtaataatagtagcgaactaaag 1438  
 Db 1229 CCGTCTATGGCGTTAATGGGCTGAAGTTTACTAATAATGCAAAACACAGCAGCAATCG 1288  
 QY 1439 gtactactcgttattacaaaagaataatggttttgcgtgataactaatgattgaggtgag 1498  
 Db 1289 GCACCTACTCGTATTACACAGATAAAATTTGGCTTTGCTCGAGATGCTGATGTTGATGAAA 1348  
 QY 1499 aaagcaaaccttattctgacacgaaaagcttaaaagttggcaacagcaccctaacaacgtg 1558  
 Db 1349 ACAAGCACCATTATTGCTATAAAACAACTTAAGTGGGTAGTGTGCAATTTACCATAG 1408  
 QY 1559 gtgattgactgttaataacaccactggtataaaacaaatcccaagtcggtgcttaatgca 1618  
 Db 1409 ACAATGGCATTTGATCAGGTAAATAAAGATCAGTAATCTTTCGCCAAMGGTAGCAGTCTA 1468  
 QY 1619 ttaaatgtgacagctgctataatggttgcaaa taccctcagcagcagtcggcactgctc 1678  
 Db 1469 AGATGCGGTTTACCATCGACAGCTCAAAAGCGCCCAAGCCCTACTTTAAACGACGCGCGTG 1528  
 QY 1679 gtattaccagaagaaaaattggttttgcgttactaaatgattgaggttgaatacaaacac 1738  
 Db 1529 GCATCAGTCTCACCTACTGAAATATCAGTTGATGCTAAGAGTGGCAATGTTACCGCCC 1588









181	ctcaatggcagtgcttatgctcaacaaattactaccagaatcgaaattggtcacaacaaac	240
QY		
181	CTCAATGGCAGTGCTTATGCT---CAAAATAATAGCAAGATCGCATTTGGTACCACAGGC	237
Db		
241	aagataaacacacgcgcgaaaggcgatgcctcagcagcaggtgaagcatccattgctttt	300
QY		
238	AA-----CAATGACAATGCCCTCGGCTAGCAATGAAGCATCCATTGGCTATT	282
Db		
301	ggtatcttttcaaggcacaaaggctctcaagctattgtctatcggtagtgcataaacccagat	360
QY		
283	GGTAGTCTTGCTAAGGCACATGCCAATCAAGCTATTGCTATCGGTGGTAGCAAAACCAAGAT	342
Db		
361	cc---taataatggtagtaatggttaatggttccccacgcgaaggtgaagcgtccatc	417
QY		
343	CCTCGTAATCAAGCGGCTTAATCAGAAGCAGGTTCCCAAGGTAAGAGTCCATC	402
Db		
418	gccatcggtgatgattggtgagggtagtgcctcgattgccatcggttagtgatgac	477
QY		
403	GCCATCGGTGGTGATGCTACTGCTGAGGGTGATGCTCGATTGCCATTGGTAGTGATGAC	462
Db		
478	ttatatcttcctaaagaattcttg-----atcgaagaatgaatttcacaaact	525
QY		
463	TTATATTTTGGTAGGAATAGCACTAACTCTAAATATCAAAATGGTCTTCTTAGCACTCTT	522
Db		
526	atctcagcctaagaattataaaaaatacaaacctcaacgcgattggtataaatcaaat	585
QY		
523	ATTCAAAACCATACAGTATTAGCCCAATACGAGACTCAA---ATGGTTCTCAGAAATAT	579
Db		
586	cgacgcacaagagacacagggcacgccagtaactgaagtgaggacatcatatcacag	645
QY		
580	AGAGGCACAGCAGCAGACAGCAGCAGCTACTGCGAGTGGAGCCATGCGATATGCAAG	639
Db		
646	ggtcatttttcaacgccttttggtacatacgcacacagctgaagctcctattccttgcca	705
QY		
640	GGTCAATTTGGCCAAACCCCTTTGGTACACGGTCNAACAGCTGAAGGCAACATTCTCTGGCA	699
Db		
706	gtaggctcttgcgcgcacacgcacacacaaatcttcaatcgctgttgggttccaatgcacaa	765
QY		
700	GTAGGTCTTACCGCCCAAGCGGAAAGGATATACATCGCTATTGGTCTTAATGACACAA	759
Db		
766	gctaacgcgtttgcagcgcacgcacattggtggaatactgttagttaatttgggttcgagcc	825
QY		
760	GCTATCAATTTATGGAGCATTAGCCCTTGGTGCAGTACTCGAGTTGATTGGATTACGGT	819
Db		
826	gttgccttagttttggtctcagatccttgataggataataataacagatgccagtcgc	885
QY		
820	ATTGCCCTAGTTATGTTCTCAGATCCTTAATAATAATAATAATAATAATAATAAAGCC	879
Db		
886	tatgtaccataggtataaacgtta---gcagaccagataaaagccacccgcagcgggtgat	942
QY		
880	TATGTACACGAAGGTAATGGGTCAACATANAATCGTCTAAAGCCACCGCAATGG---	935
Db		
943	tctacggatattttcccatcttggttaatagtaataataataacagatcagcagcgtataa	1002
QY		
936	-----TTTATTTTCATTGG-----TAGTAGCACTATCAAGCGGTAAA	972
Db		
1003	atcatcaatgctcgtgcgggttctcgggataccagtcggtcgaatggtgcacagccttaaa	1062
QY		
973	ATCATCAATGCTGCTGACGTTATGAGGATACCGATGCGGTCAATGTGGCACAGCTAATA	1032
Db		
1063	ttgtggagggaactgcgtcaactcgtataaattacttttaagggtgatggtgcataataatgc	1122
QY		
1033	CGGTGGAGAACTCGCTAAGCGTCAAAATTTACTTTTAAGGGTGTATG-----ATAACGGT	1086
Db		
1123	aatagcgtagaaaaagggtttgggcaataacttttaactattaagggtgatg-----cacag	1176
QY		
1087	ACTGGCGTTAAGAAAAAATCGGGCGAGACTTTTAACCATTTAAAGCTGGTGAGACCCACGCG	1146
Db		
1177	accaagcgtataaccgaagactaacatcgcgtgtggttaacagatggtgcaa---tggctcgtaaa	1233
QY		
1147	GACAGCTAACCGATAATAATAAACATTGGTGTGGTAACAGATAATAATAATCTGCTGTGAAA	1206
Db		
1234	gttaaaacttcttaagagcgtgactggtgattgacacgtctcgcgtaccacacaaatacacc	1293
QY		

2698

ACGGTTAAAGCTTTGCTGGACAAACGGCGTTGGTGGGCTCCGTTGGGCTCAGGT

2757

5785

gctgaacgcgctatccaaaaatggtggcagcaggtgaggtcaatgcccacacacagcgtgcg

5844

2758

GCTGAACCGCTATTCAAAATGTGGCAGCAGGTGAGGTTCAGTGGCCACGACCGCATGGC

2817

5845

gtcaatgtagccagttgtacaaagcccccacaaagcattgccaacgcgaaccaatgagctt

5904

2818

GTCAATGGTAGCCAGTTGTACAAAGCCACCCAAAGCATTTGCCAAGCAACCAATGAGCTT

2877

5905

gaccatcgtatccacccaaaacgaaataaagccaatgagggatgattcaatcagcagtcgcg

5964

2878

GACCATCGTATCCACCAAAACGAAATTAAGCCAAATGAGGATTTTCATCAGCGATGGC

2937

5965

atggcgtccatcccaagcctacattcctgcagatccatggttaccgggggtattgcc

6024

2938

ATGGCGTCCATGCCACAAGCCTACATCTCTGCACATCCATGCCAAGCCATGAGTATGGC

2997

6025

accacaacggttcaaggtggtggtgagctggagctgtcgaaagctgtcgataatggtcaa

6084

2998

ACCCAACCGTCAAGGTGGGTGGCAGTGGGAGTGTCTGAAAGCTGTGCGGATTAATGCTCAA

3057

6085

tggttataaataaatggttcagccgatacccaagggccatgtagggcgcaggttgtt

6144

3058

TGGGTATTTAAATCAATGTTTCAGCCGATACCCAAGCCCATGATAGGCGCGCAGTTGCT

3117

6145

gcaggttttcaactt 6159

3118

GCAGTTTTCACTTT 3132

RESULT 10

AX079922

LOCUS

DEFINITION

SEQUENCE

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 6942)

Loomsore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.

Recombinant high molecular weight major outer membrane protein of moraxella

CONNAUGHT LABORATORIES LIMITED (CA)

Location/Qualifiers

1. .6942

/organism="Moraxella catarrhalis"

/db\_xref="taxon:480"

2286 a 1621 c 1564 g 1471 t

22-FEB-2001

PAT

22-FEB-2001

6942 bp

DNA

10 from Patent WO0107619.

AX079922

AX079922.1

GI:13159441

Moraxella catarrhalis.

Moraxella catarrhalis.

Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Mismatches

<





QY	3382	ggcaagatttatgatttaaataacccgaacttgaataataatcagaagctgcgcaaaaca	3441
Db	3184	GGTCGGGTTTTATGATTATAAAACCGCAACTTGAAGCAGAAAATCAACAGTGCTGTATAACA	3243
QY	3442	gcacaaacttaatacacgaattctcagttagcatgaacaaaggtaataactttaaggtt	3501
Db	3244	GCACAAAACATCATACACGAATTTCTAGTAGCAGATGACAAAGGTANTCACTTTACGGTT	3303
QY	3502	agtaaccttactccagttatgacacctcaaaagacccttgatgcatcacctttgcaagt	3561
Db	3304	AGTAACCCCTTACTCCAGTTATGACACCTCAAAGACCCTCTGATGTTCATCACCTTTGCAAGGT	3363
QY	3562	gaaaacggtatcacaccaaggtataaataaggttgtgtggtggtggttcattgaccaaacc	3621
Db	3364	GAAAAAGGCATTTACCAACCAAGGTAAATAAAGGTGTGGTGGCTGTGGGCATTTGACCAAAACC	3423
QY	3622	aaagcttaaccacgcttaagctgacogtgggttaataataatggtcaaaaggcattgtcatt	3681
Db	3424	AAAGCCTTTAACACGCCCTTAACCTGACCGTGGGTAAATAATAATGTCAAAAGGCATTTGTCAATT	3483
QY	3682	aacgccaaaaatggttcaaaaataccatcacaggactaacgaacactctagclaatgttac	3741
Db	3484	GACAGTAAGAATGTTCAAAATACCATCACAGGACTTAAGCAACACTCTAGCTTAATGTTTACC	3543
QY	3742	aatgataaagtagcgtagcagcacagacagggcaataataacaaagacgaagacaaa	3801
Db	3544	AATGA-----TGGTGCAGGACACGCCACTTAGCCNAAGGGCTTGCCAATGACACCGACAAA	3597
QY	3802	accgtgcccgcagcattgtgatgtgaagcgcaggctttaacttcgaagggaattggt	3861
Db	3598	ACCCGTGCCGCCAGCATTTGGTGATGTGCTTAACAGCAGCGCTTTAACTTCAAGGGCAATGGT	3657
QY	3862	gaagcggttgactttgtctcacttatcaacgctcaactttgccaattgccaatgccaataccacc	3921
Db	3658	GAAAGCGTTGACTTTGTCTCCACTTATGACACTGTTTGACTTTATCGATGGCAATGCCACC	3717
QY	3922	accgtctaaggtgaccttatgatgacacaaagcaaaacagtaagttggtctatgatcgaat	3981
Db	3718	ACCGCTAAGGTGACCTATGNTGACACAAAGCAAAACAGTAAGTGGTCTATGATGTCAAT	3777
QY	3982	tgtgatgatacaacattgaattaaa---gataaaaacttggcgtataaaacccccacca	4038
Db	3778	GTGGATTAATAAAACCACTTGAAGTGACAGTGTATAAAAAACTTTGGCTCAAAACCAACACA	3837
QY	4039	tgaccagtlactggcacaggtgctataaatttgccttaagcaataoaagctactggcat	4098
Db	3838	CTGACCAAAAACAGTGTCTAATGGTAATGCA--ACCAAAATTTAGTGCCCGCCGATGCGCAT	3894
QY	4099	gcctgtgtaagcgaagtatatcgtgtcattctataacacaccttatctggcacatccaa	4158
Db	3895	GCCTTTGTTAAAGCCAGTATATCGCCACCCTCTAATACCTTGGCTGGCGACATCCAA	3954
QY	4159	actgcaaaagggaagcgaagcaaacactgcagcaggctatgtgtagctgatggcaat	4218
Db	3955	ACCGCCAAAGGGCAAGCCAAAGCAGCTCAGCAAGCTATGTGGATGTGATGGCAAC	4014
QY	4219	aaggtcatctatgacagtaacgatacaagactactatcaagcgaataatgatgcagitt	4278
Db	4015	AAGTCACTCTATGACAGTACCGATAAGAAAGTACTATCAAGTCAATGACAAAGGTCAAGTG	4074
QY	4279	gataaaacaaaagttgcgaagcaaaactggtgcgaagcccccaacccccagatggc	4338
Db	4075	GACAAAACAAAGNAGTTGCCAAAGCAAACTGGTGC GCCCAAGCCCCCAAGATGGC	4134
QY	4339	acattggtcnaatgtaatgtaaatcagtcattataacaaaagaacaaagtaaatgacaaat	4398
Db	4135	ACATTTGGCTCAAATGTAATGTAATCATAGTCTATTAAACAAAGCAAGTAATGATGCCAAT	4194
QY	4399	aaaagcaaggcatcaatgaagcaaacgctttgtttaaaggacttgaagaagcgccttct	4458
Db	4195	AAAAGCAAGGCATCAATGAAGCAACACGCCCTTTATCAAAAGGGCTTTGAAAACGCCGCCAAA	4254
QY	4459	gatacaaaaacaaaaacccgagtaactgtgggtgattttaaatgcogttgcccnaaaca	4518
Db	4255	GACACAAAACCAAAAACCGCAGTAACCTGTGGGTGATTAAATGCCGTGCCCAANACA	4314
QY	4519	cgcgtacotcttgcaggggatacagcacaacggtctaaaaaactggcgagactttgacc	4578
Db	4315	CCCTCACCCTTTGCAGGGGATACAGGCACAACGGCTTAAAAAACTGGCCGAGACTTTTGACC	4374
QY	4579	atcaaaaggltggcacaacacacccaataagctaaacccgataataacatcctcgtgtggtagca	4638
Db	4375	ATCAAAAGGTGGCAACACAGACCAATAAGCTAACCGATAATAACATCTGGTGTGGTAGCA	4434
QY	4639	ggtactgatggcttcactctgcaaaacttgcaaaagacctaaccaatcttaacagcgttaat	4698
Db	4435	GGTATGATGGCTTCACTGTCAAACTTGCCAAAGACTTAACCAATCTTAAACACGGTTAAT	4494
QY	4699	gcaggtggcaccacaaattgatgaaaaggcatctctttgtagacgcaaacggtcgaagcc	4758
Db	4495	GCAGGTGGCACCAAGAAATTGATGAAAAGGCATCTCTTTTGTAGACGCAAAACGGTCAAGCC	4554
QY	4759	aaagcaaacacccctgtgctaaagtgcacatg99gctggagacctgggtggc-----	4806
Db	4555	AAAGCAAAACCCCTGTGCTAAAGTGCCAAATGGGCTGGACCTGGGTGGCAAAACGCATCAGT	4614
QY	4807	-----	4806
Db	4615	AACATCGTGCAGCTGTTGATGATAACGATGCGGTGAACCTTAAAGCAGTTTAAATGAAGTT	4674
QY	4807	-----	4806
Db	4675	GCCAAAACGGTCAACAACCTAAACAACAAAAGTAACCTCAGGTGGGTCTATTACCCCTTTTGTG	4734
QY	4807	-----	4806
Db	4735	GTAACCGATGCCAATGGCAAGCCCATCAATGGCACCGGATGGCAAGCCCAAAAGCCATC	4794
QY	4807	-----	4806
Db	4795	AAGGCGCGATGTTAAATACTATCACGCCAAAGCCAAACGGCGTACCTGTGGACAAAGAT	4854
QY	4807	-----	4806
Db	4855	GGCAAGCCCATCACCGATGCGGCAAAACTTGCCAACTCTGGCAGTCTATGGCAAAACCCCTT	4914
QY	4807	-----	4806
Db	4915	GATGAGGTTCATCAAGTGGTGGCAAGCCTAGGCGGCAACTCAGATGCCATCACCCCTAACC	4974
QY	4807	-----	4806
Db	4975	AAACATCAAGTCCACTTTGCCACAAATTTGACACNCCAAACACANGTAAATGCCAATGCAGGG	5034
QY	4807	-----	4806
Db	5035	CAAGCCCAAGTCTGCCACCGCTATCAGCAGCACAGCAAAAGTAATGCTGCCAGTGTCAAA	5094
QY	4807	-----	4806
Db	5095	GATGTCTAAATGTAGGCTTTAACTTGCAGACCAATCAACATCAAGTGACTTTGTCAAA	5154
QY	4807	-----	4806
Db	5155	GCCTATGATACCGTCAACTTTGTCAATGTTACAGGTGGCGACATCACAAAGCGTGCCTAGT	5214
QY	4807	-----	4806
Db	5215	GCATGACGACGATGAGTAACATCACCGTCAACACCGCCTTAGCAGCGACCGATGATGAT	5274



[illegible]

```

QY 5365 ataggcagacaaacccaagcaggaacccaatccatcgccatcggtgataacgcacagcc 5424
|||||
Db 2338 ATAGGCAGACAAACCCAAAGCAGCAACCAATCCATCCGTCGTGATACGCACAGCC 2397
|||||
QY 5425 acggcgatcaatccatcgccatcggtacagcgaatgtgtagcaggtgaag 5475
|||||
Db 2398 ACGGGCATCAATCCATCCGTCGTGATACGCAGCAATGTGTAGCAGGTAAG 2448
|||||

RESULT 12
AC007926
LOCUS 149789 bp DNA HTG 12-FEB-2001
DEFINITION Trypanosoma brucei chromosome II clone RCI193-3H15, *** SEQUENCING
IN PROGRESS ***, 1 ordered pieces.
AC007926
VERSION AC007926.7 GI:12746529
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE
1 (bases 1 to 149789)
El-Sayed, N., Zhao, H., Mason, T., Wanless, D., Fujii, C., Barnstead, M.,
Bowman, C., Peterson, J., Khalak, H., Ullu, E., Melville, S.,
Donelson, J., White, O., Fraser, C. and Adams, M.
Trypanosoma brucei GUTat10.1 RCI193-3H15 BAC genomic sequence
Unpublished
2 (bases 1 to 149789)
El-Sayed, N.M., Khalak, H. and Adams, M.D.
Direct Submission
Submitted (28-JUN-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Feb 12, 2001 this sequence version replaced gi:6707792.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 149789: contig of 149789 bp in length.

FEATURES
Source
1..149789
/organism="Trypanosoma brucei"
/isolate="GUTat10.1"
/db_xref="taxon:5691"
/chromosome="II"
/clone="RCI193-3H15"

BASE COUNT 41249 a 31075 c 30327 g 47138 t
ORIGIN

Query Match 1.7%; Score 103; DB 60; Length 149789;
Best Local Similarity 46.5%; Pred. No. 5e-12;
Matches 334; Conservative 0; Mismatches 385; Indels 0; Gaps 0;

QY 1398 atggaattactaatgataagtagtagcgaactaaaggtagctactgtattaccac 1457
|||||
Db 6980 ATACGATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 69939
|||||
QY 1458 aagaaaaattggttctgctggtactaatgatggagttgatgaagcaaaccttatcttga 1517
|||||
Db 69940 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 69999
|||||
QY 1518 caacgaagctaaaagttggcaacagcaccctaaacagtggttagctgttaataa 1577
|||||
Db 70000 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 70059
|||||
QY 1578 caccactggaataaacaacatcccaagtcggtgcttaatggcattaaatttgcacagtcgc 1637
|||||

```

```

Db 70060 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 70119
|||||
QY 1638 taataattgttgcaaatccctcagcaacagtcggcactgctcgctattaccggagagaaaa 1697
|||||
Db 70120 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 70179
|||||
QY 1698 tggtttgcgtggtactaatgatggagttgatgaacagcaccacatttgggataaagaacg 1757
|||||
Db 70180 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 70239
|||||
QY 1758 actaaagtggttcgctgttgaaattaccacagatagtggttattactgctgtaataccaa 1817
|||||
Db 70240 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 70299
|||||
QY 1818 gattaccggacttactaatggtatagcaaataccgctgcggttaccatcaaacagctcaa 1877
|||||
Db 70300 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 70359
|||||
QY 1878 agacgcaacgcttactttaaacgcagcgatggcatcagtagtattaaatagtaacgggga 1937
|||||
Db 70360 TGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 70419
|||||
QY 1938 tctagttgatagtagtgccaattaccaccccaactataacattagcgtggaaccac 1997
|||||
Db 70420 TAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 70479
|||||
QY 1998 taagctttaacagtaatggcaccagtggttaataataaatttagtggtagtaagctcaga 2057
|||||
Db 70480 TGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 70539
|||||
QY 2058 taacaatagcttggttaccgccaaagattggcagagctatctataaagtaacatgaaa 2116
|||||
Db 70540 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 70598
|||||

RESULT 13
PM0277635
LOCUS 3900 bp DNA BCT 21-OCT-2000
DEFINITION Pasteurella multocida mapA gene.
ACCESSION AJ277635
VERSION AJ277635.1 GI:10945096
KEYWORDS MapA protein.
SOURCE Pasteurella multocida.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Pasteurella.
REFERENCE
1 (bases 1 to 3900)
Henderson, I.R., Nataro, J.P., Cappello, R. and Stein, C.
Evolutionary origins of the autotransporter proteins
Unpublished
REFERENCE
2 (bases 1 to 3900)
Henderson, I.R.
Direct Submission
Submitted (25-APR-2000) Henderson I.R., Center for Vaccine
Development, University of Maryland School of Medicine, 685 W.
Baltimore St, MD 21202, USA
FEATURES
Location/Qualifiers
source 1..3900
/organism="Pasteurella multocida"
/strain="PM70"
/db_xref="taxon:747"
gene 1..3900
/cds 1..3900
/genes="mapA"
/function="putative autotransporter protein"
/codon_start=1
/product="MapA protein"
/protein_id="CAC14202.1"
/db_xref="GI:10945097"
/translation="MKRIYKTLWNAATQSVVVVSELAKAGGKSASKSALVNSVSGFS
FTLIAASVVLGGQVNAAEVTTGNTGVSGDDKYCFYNASSQSVICGDTATTKTDKDNK

```







Thu Sep 13 14:19:03 2001

us-09-361-619-8.rge

```
Best Local Similarity 50.0%; Pred. No. 4.7e-08;
Matches 245; Conservative 0; Mismatches 236; Indels 9; Gaps 1;

QY 5677 aaatctgacgacagcaggtacacacacacagcaggtgcccacaggtacggttaaaagc 5736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2803 AAAGCTGACGGTACTGCGGATAAAACCAAGCGAAGTGACGAATGATAAAGTTTCTACC 2862

QY 5737 ttgtctggacaaacagcggttgggtgcgtccgtgggtgcctcaggtgctgaacgccgt 5796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2863 GATGAAAAACACGCTTGTCTACGCTTGATCCAAATGATCAATCAAAAAGGTAAGGTCTCGTG 2922

QY 5797 atccaaaatgtggcagcaggtgaggtcagtcacacacagcagcagcagtcggtcaatggtagc 5856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2923 ATTGACAATGTGGCTAATGGCGATATTTCTGCCACTTCACCCGATGCCGATTAACCGAAGT 2982

QY 5857 cagttgtac-----aaagcaccacaaagcattgccacgcgaaccaatgagcttgac 5907
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2983 CAGTTGTATGCTGTGGCAAAAGGGTAACAACCTTTGCTGGACAAAGTGAATAATCTTGAG 3042

QY 5908 catcgtatccacacaaacaaataaaagcaatgagggatttcacagcagtgatggcgatg 5967
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3043 GGCAAAAGTGAATAAAGTGGCAAAACGTGCAGATGCAGGTACAGCAAGTGCATTAGCGGCT 3102

QY 5968 gcgtccatgccaaagcctacattcctggcagatccatggttaccgggggtattgccacc 6027
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3103 TCACAGTTACCAACAAGCCACTATGCCAGGTAATCAATGTTGCTATTGCGGGAAGTAGT 3162

QY 6028 cacaacggtcaaggtgcggtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtg 6087
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3163 TATCAAGGTCAAAATCGTTTAGCTATCGGGGTATCAAGAAATTTCCGATATATGGCAAAAGTG 3222

QY 6088 gtatttaaatcaatggttcagcagcagcagcagcagcagcagcagcagcagcagcagcag 6147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3223 ATTATTCGCTTGTCCAGCACACCAATAGTCAAGGTAAACAGGCGTTGCAGCAGGTGTT 3282

QY 6148 ggttttcaact 6157
    ||| ||| ||| |||
Db 3283 GGTACCAGT 3292
```

Search completed: September 12, 2001, 23:04:48  
Job time: 37510 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 13:29:49 ; Search time 1369.14 Seconds  
(without alignments)  
2824.580 Million cell updates/sec

Title: US-09-361-619-8

Perfect score: 6159

Sequence: 1 atgaatcacattataaagt.....ttggtcagggttttcacttt 6159

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601.\*

```

1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6159	100.0	6159	22	AAF59104
2	4101.6	66.6	6144	22	AAF59103
3	4101.6	66.6	6972	22	AAF59102
4	4090.6	66.4	6971	17	AAT38740
5	4090.6	66.4	6973	22	AAF59100
6	3982	64.7	62909	22	AAF28545
7	3947.8	64.1	5979	22	AAF59101
8	3946.8	64.1	6259	22	AAF59106
9	2939.2	47.7	6942	22	AAF59105
10	166.2	2.7	720	22	AAF59129
11	144.6	2.3	3030	21	AAA92496

12	143	2.3	3036	21	AAA92493	Haemophilus influe
13	138.2	2.2	3354	21	AAA92495	Haemophilus influe
14	129.4	2.1	3342	21	AAA92498	Haemophilus influe
C 15	122	2.0	936	22	AAF58252	Oligonucleotide D1
C 16	122	2.0	936	22	AAF58254	Oligonucleotide D1
C 17	122	2.0	936	22	AAF58257	Oligonucleotide D1
C 18	122	2.0	936	22	AAF58259	Oligonucleotide D2
C 19	122	2.0	936	22	AAF58262	Oligonucleotide D2
C 20	122	2.0	936	22	AAF58255	Oligonucleotide D1
21	119.2	1.9	936	22	AAF58252	Oligonucleotide D1
22	119.2	1.9	936	22	AAF58254	Oligonucleotide D1
23	119.2	1.9	936	22	AAF58257	Oligonucleotide D1
24	119.2	1.9	936	22	AAF58259	Oligonucleotide D2
25	119.2	1.9	936	22	AAF58262	Oligonucleotide D2
26	119.2	1.9	938	22	AAF58255	Oligonucleotide D1
27	84.4	1.4	3294	17	AAT41475	Haemophilus adhesi
28	83.6	1.4	2037	17	AAT41477	Haemophilus adhesi
29	82	1.3	2079	21	AAA92494	Haemophilus influe
30	81.8	1.3	7486	19	AAV22837	Haemophilus paraga
31	80.2	1.3	8930	19	AAV22834	Haemophilus influe
32	77.2	1.3	7253	21	AAA92499	Haemophilus influe
33	77.2	1.3	7291	17	AAT41476	Haemophilus adhesi
34	76.4	1.2	3300	21	AAA92497	Haemophilus influe
C 35	75	1.2	1830121	17	AAV42063	Haemophilus influe
36	65.2	1.1	1757	21	AAA92500	Haemophilus influe
37	60	1.0	60	17	AAT38741	Moraxella outer me
38	58.4	0.9	1797	20	AAV85794	DNA encoding a sur
39	56.8	0.9	1797	20	AAV85790	DNA encoding a sur
40	55.8	0.9	1779	20	AAV99125	DNA encoding N. me
41	55.4	0.9	3567	21	AAV70117	Plasmodium falcipa
42	55.2	0.9	1770	20	AAV85797	DNA encoding a sur
43	55.2	0.9	1776	20	AAV99124	DNA encoding N. me
44	55.2	0.9	1776	20	AAV85798	DNA encoding a sur
45	55.2	0.9	1776	20	AAV85793	DNA encoding a sur

#### ALIGNMENTS

RESULT 1

AAF59104  
ID AAF59104 standard; DNA; 6159 BP.

XX AAF59104;

AC AAF59104;

DT 24-APR-2001 (first entry)

XX M. catarrhalis strain Q8 200kDa gene SEQ ID NO:8.

DE Moraxella catarrhalis strain Q8; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KW otitis media; detection; ds.

XX Moraxella catarrhalis.

OS Moraxella catarrhalis.

XX WO200107619-Al.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

PI Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

DR P-PSDB; AAB69135.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

PT useful in protective vaccines and for diagnosis

XX

Claim 1: Fig 4A-V; 247pp; English.

xx The present invention describes an isolated and purified nucleic acid (I)  
cc that encodes a 200 kDa outer membrane protein of *Moraxella catarrhalis*.  
cc The 200 kDa outer membrane protein (II) has antibacterial activity and  
cc can be used in vaccines. (II), and its truncated versions, are used as  
cc immunogenic compositions and vaccines to protect against *M. catarrhalis*  
cc infections, particularly otitis media in humans. (II) is also used as  
cc antigen in immunoassays for detecting specific antibodies (Ab), and to  
cc generate Ab. (I) are used for recombinant production of (II) and its  
cc fragments are used as probes for identifying/cloning 200 kDa protein  
cc genes from other strains, and for diagnostic detection of *M. catarrhalis*.  
cc (I) makes possible production of large amount of recombinant immunogens.  
cc Expression of truncated versions of (II) reduces toxicity of the protein  
cc towards the *Escherichia coli* host. The present sequence represents the  
cc *M. catarrhalis* strain Q8 200kDa gene, which is given in the  
cc identification of the present invention.

XX	2035 A.	2035 B.	2035 C.	2035 D.	2035 E.	2035 F.	2035 G.	2035 H.	2035 I.	2035 J.	2035 K.	2035 L.	2035 M.	2035 N.	2035 O.	2035 P.	2035 Q.	2035 R.	2035 S.	2035 T.	2035 U.	2035 V.	2035 W.	2035 X.	2035 Y.	2035 Z.	2035 AA.	2035 AB.	2035 AC.	2035 AD.	2035 AE.	2035 AF.	2035 AG.	2035 AH.	2035 AI.	2035 AJ.	2035 AK.	2035 AL.	2035 AM.	2035 AN.	2035 AO.	2035 AP.	2035 AQ.	2035 AR.	2035 AS.	2035 AT.	2035 AU.	2035 AV.	2035 AW.	2035 AX.	2035 AY.	2035 AZ.	2035 BA.	2035 BB.	2035 BC.	2035 BD.	2035 BE.	2035 BF.	2035 BG.	2035 BH.	2035 BI.	2035 BJ.	2035 BK.	2035 BL.	2035 BM.	2035 BN.	2035 BO.	2035 BP.	2035 BQ.	2035 BR.	2035 BS.	2035 BT.	2035 BU.	2035 BV.	2035 BW.	2035 BX.	2035 BY.	2035 BZ.	2035 CA.	2035 CB.	2035 CC.	2035 CD.	2035 CE.	2035 CF.	2035 CG.	2035 CH.	2035 CI.	2035 CJ.	2035 CK.	2035 CL.	2035 CM.	2035 CN.	2035 CO.	2035 CP.	2035 CQ.	2035 CR.	2035 CS.	2035 CT.	2035 CU.	2035 CV.	2035 CW.	2035 CX.	2035 CY.	2035 CZ.	2035 DA.	2035 DB.	2035 DC.	2035 DD.	2035 DE.	2035 DF.	2035 DG.	2035 DH.	2035 DI.	2035 DJ.	2035 DK.	2035 DL.	2035 DM.	2035 DN.	2035 DO.	2035 DP.	2035 DQ.	2035 DR.	2035 DS.	2035 DT.	2035 DU.	2035 DV.	2035 DW.	2035 DX.	2035 DY.	2035 DZ.	2035 EA.	2035 EB.	2035 EC.	2035 ED.	2035 EE.	2035 EF.	2035 EG.	2035 EH.	2035 EI.	2035 EJ.	2035 EK.	2035 EL.	2035 EM.	2035 EN.	2035 EO.	2035 EP.	2035 EQ.	2035 ER.	2035 ES.	2035 ET.	2035 EU.	2035 EV.	2035 EW.	2035 EX.	2035 EY.	2035 EZ.	2035 FA.	2035 FB.	2035 FC.	2035 FD.	2035 FE.	2035 FF.	2035 FG.	2035 FH.	2035 FI.	2035 FJ.	2035 FK.	2035 FL.	2035 FM.	2035 FN.	2035 FO.	2035 FP.	2035 FQ.	2035 FR.	2035 FS.	2035 FT.	2035 FU.	2035 FV.	2035 FW.	2035 FX.	2035 FY.	2035 FZ.	2035 GA.	2035 GB.	2035 GC.	2035 GD.	2035 GE.	2035 GF.	2035 GG.	2035 GH.	2035 GI.	2035 GJ.	2035 GK.	2035 GL.	2035 GM.	2035 GN.	2035 GO.	2035 GP.	2035 GQ.	2035 GR.	2035 GS.	2035 GT.	2035 GU.	2035 GV.	2035 GW.	2035 GX.	2035 GY.	2035 GZ.	2035 HA.	2035 HB.	2035 HC.	2035 HD.	2035 HE.	2035 HF.	2035 HG.	2035 HH.	2035 HI.	2035 HJ.	2035 HK.	2035 HL.	2035 HM.	2035 HN.	2035 HO.	2035 HP.	2035 HQ.	2035 HR.	2035 HS.	2035 HT.	2035 HU.	2035 HV.	2035 HW.	2035 HX.	2035 HY.	2035 HZ.	2035 IA.	2035 IB.	2035 IC.	2035 ID.	2035 IE.	2035 IF.	2035 IG.	2035 IH.	2035 II.	2035 IJ.	2035 IK.	2035 IL.	2035 IM.	2035 IN.	2035 IO.	2035 IP.	2035 IQ.	2035 IR.	2035 IS.	2035 IT.	2035 IU.	2035 IV.	2035 IW.	2035 IX.	2035 IY.	2035 IZ.	2035 JA.	2035 JB.	2035 JC.	2035 JD.	2035 JE.	2035 JF.	2035 JG.	2035 JH.	2035 JI.	2035 JJ.	2035 JK.	2035 JL.	2035 JM.	2035 JN.	2035 JO.	2035 JP.	2035 JQ.	2035 JR.	2035 JS.	2035 JT.	2035 JU.	2035 JV.	2035 JW.	2035 JX.	2035 JY.	2035 JZ.	2035 KA.	2035 KB.	2035 KC.	2035 KD.	2035 KE.	2035 KF.	2035 KG.	2035 KH.	2035 KI.	2035 KJ.	2035 KK.	2035 KL.	2035 KM.	2035 KN.	2035 KO.	2035 KP.	2035 KQ.	2035 KR.	2035 KS.	2035 KT.	2035 KU.	2035 KV.	2035 KW.	2035 KX.	2035 KY.	2035 KZ.	2035 LA.	2035 LB.	2035 LC.	2035 LD.	2035 LE.	2035 LF.	2035 LG.	2035 LH.	2035 LI.	2035 LJ.	2035 LK.	2035 LL.	2035 LM.	2035 LN.	2035 LO.	2035 LP.	2035 LQ.	2035 LR.	2035 LS.	2035 LT.	2035 LU.	2035 LV.	2035 LW.	2035 LX.	2035 LY.	2035 LZ.	2035 MA.	2035 MB.	2035 MC.	2035 MD.	2035 ME.	2035 MF.	2035 MG.	2035 MH.	2035 MI.	2035 MJ.	2035 MK.	2035 ML.	2035 MM.	2035 MN.	2035 MO.	2035 MP.	2035 MQ.	2035 MR.	2035 MS.	2035 MT.	2035 MU.	2035 MV.	2035 MW.	2035 MX.	2035 MY.	2035 MZ.	2035 NA.	2035 NB.	2035 NC.	20
----	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----

Query Match	100.0%;	Score 6159;	DB 22;	Length 6159;
Best Local Similarity	100.0%;	Pred. No. 0;		
		Indels 0;	Gaps 0;	

QY	1	atgaatcacatctataaagtcattctttaacaaagccacagacattttaggcgtggcg	60
Db	1	atgaaatcacatctataaagtcattctttaacaaagccacagacattttaggcgtggcg	60
QY	61	gaatatgccaaatcccacagtagcgggggggaactgtctacagggcaagttagcgagt	120
Db	61	gaatatgccaaatcccacagtagcgggggggaactgtctacagggcaagttagcgagt	120
QY	121	gtacgcactctaaagctttggccgtattggcgcgcgtcgtcttcctcgtatcgatcgagcg	180
Db	121	gtacgcactctaaagctttggccgtattggcgcgcgtcgtcttcctcgtatcgatcgagcg	180
QY	181	ctcaatggcagtgcttatgtctcaaaaattactaccaagatcgaaaattggtcaacaac	240
Db	181	ctcaatggcagtgcttatgtctcaaaaattactaccaagatcgaaaattggtcaacaac	240
QY	241	aagataaacacacgcgtgaaaggcgatgccctagcagacaagtgaaagcatccattgctttt	300
Db	241	aagataaacacacgcgtgaaaggcgatgccctagcagacaagtgaaagcatccattgctttt	300
QY	301	ggtagctttctaaagccaaaggctctcaagctattgctatcgtagtgcattggtcaaacccagat	360
Db	301	ggtagctttctaaagccaaaggctctcaagctattgctatcgtagtgcattggtcaaacccagat	360
QY	361	cctaataatggtagtaattggttaattgtaggtttccacgcacaaaggttaacgagttccatgcgc	420
Db	361	cctaataatggtagtaattggttaattgtaggtttccacgcacaaaggttaacgagttccatgcgc	420
QY	421	atcggtaggtgatgtattggctgaagggtgatgcctcgattgcacatcggtagtgatgactta	480
Db	421	atcggtaggtgatgtattggctgaagggtgatgcctcgattgcacatcggtagtgatgactta	480
QY	481	tatttgcttaagaattcttgatctgaagaatgaatttcacaaacttattcatggccatgaa	540
Db	481	tatttgcttaagaattcttgatctgaagaatgaatttcacaaacttattcatggccatgaa	540
QY	541	ataattaaaaaaatacaaacctccaccgatggttaaaatacaaatatcgacgcacaagaaca	600
Db	541	ataattaaaaaaatacaaacctccaccgatggttaaaatacaaatatcgacgcacaagaaca	600
QY	601	caaggccacccagtagctcagttggaggccatgtcatatgcacagggctcattttcccaac	660
Db	601	caaggccacccagtagctcagttggaggccatgtcatatgcacagggctcattttcccaac	660
QY	661	gcctttggtatatacgcacacagctgaagctgcttatctccttgacagtctcttgcgcgc	720
Db	661	gcctttggtatatacgcacacagctgaagctgcttatctccttgacagtctcttgcgcgc	720





## RESULT 2

AA59103  
 ID AAF59103 standard; DNA; 6144 BP.  
 AC  
 AC AAF59103;  
 XX  
 XX  
 DT 24-APR-2001 (first entry)  
 XX  
 DE M. catarrhalis strain 4223 genomic 200kDa coding sequence SEQ ID NO:6.  
 XX  
 KW Moraxella catarrhalis strain 4223; major outer membrane protein;  
 KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
 KW otitis media; detection; ds.  
 XX  
 OS Moraxella catarrhalis.  
 XX  
 XX WO200107619-A1.  
 PN  
 XX  
 PD 01-FEB-2001.  
 XX  
 XX 26-JUL-2000; 2000WO-CA00870.  
 XX  
 XX 27-JUL-1999; 99US-0361619.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Loosmore SM, Sasaki K, Yang Y, Klein MH;  
 XX  
 XX WPI: 2001-159722/16.  
 DR  
 DR P-PSDB; AAB69134.  
 XX  
 XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
 PT useful in protective vaccines and for diagnosis  
 XX  
 XX Claim 1; Fig 3A-W; 247pp; English.  
 XX  
 CC The present invention describes an isolated and purified nucleic acid (I)  
 CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.  
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and  
 CC can be used in vaccines. (II), and its truncated versions, are used as  
 CC immunogenic compositions and vaccines to protect against M. catarrhalis  
 CC infections, particularly otitis media in humans. (II) is also used as  
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to  
 CC generate Ab. (I) are used for recombinant production of (II), and its  
 CC fragments are used as probes for identifying/cloning 200 kDa protein  
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
 CC (I) makes possible production of large amount of recombinant immunogens.  
 CC Expression of truncated versions of (II) reduces toxicity of the protein  
 CC towards the Escherichia coli host. The present sequence represents the  
 CC M. catarrhalis strain 4223 genomic 200kDa coding sequence, which is  
 CC given in the exemplification of the present invention.  
 XX  
 XX Sequence 6144 BP; 2014 A; 1403 C; 1399 G; 1328 T; 0 other;  
 SQ

Query Match 66.6%; Score 4101.6; DB 22; Length 6144;  
 Best Local Similarity 81.3%; Pred. NO. 0;  
 Matches 5063; Conservative 0; Mismatches 1009; Indels 156; Gaps 20;

QY 1 atgaatcacatctataaagtcattctttaacaaagccacagggcacatttatggcgtggcg 60  
 DB 1 atgaatcacatctataaagtcattctttaacaaagccacagggcacatttatggcgtggcg 60

QY 61 gaatatgccaaatccacagtcaggggggtgagctgtgctacagggcgaagtggcag 120  
 DB 61 gatlacgcaaatccacagtcaggggggtgagctgtgctacagggcgaagtggcag 120

QY 121 gtacgcacttaagctttgccctgattgcgcgctgctgctcctgctgagtcggcgac 180  
 DB 121 gtatgcacttgagctttgccctgattgcgcgctgctgctcctgctgagtcggcgac 180

QY 181 ctcaatggcagctcttatgctcaacaattactacca---agatcgaaattggtcaaca 237  
 DB 181 ctcaatggcagctcttatgctcaacaattactacca---agatcgaaattggtcaaca 237

DB 181 ctcaatggcagctcttatgctcaacaattactaccaattatcgcaattggtgaacaa 240  
 QY 238 acaagataaaacacacgcgtgaagggcgatgccctagtcgacaggtggaagcatccattgct 297  
 DB 241 aaccagccaagacgctc-----agcactgccaagggcgagctgacgagcattgct 294  
 QY 298 ttggtagtctttcttaaggcacaaggctctcaagctattgctatcggtagtgctcaacca 357  
 DB 295 attggtgaaatgctaaacgcacagggcggtcaagccatcgccatcggtagtagtaataa 354  
 QY 358 gatcctaataatgtagtaataatgtaattgagttccacaccccaaggaagtaacgaatc 417  
 DB 355 actgtcaatggaagcagatttggaataagatagggtaccgatgctacgggtcaagagatcc 414  
 QY 418 gccatcggtggtgatgtattggctgaggggtgatgccctcgattgccatcggtagtgatgac 477  
 DB 415 gccatcggtggtgatgtataaggctagtggtgatgctcgattgccatcggtagtgatgac 474  
 QY 478 ttatttgcctaagaattctgactctgaagaa-----tgaatttcaacaaa 522  
 DB 475 ttacatttgcctgactgacgtggttaactctaaacatccgaaaggctactctgtattacgat 534  
 QY 523 ctatttcattgccaataattataaaaaatacaaacctcaaccgatggtataaatcaaa 582  
 DB 535 cttattacggccatgacgtattataaagaataacgagctcaaggataatgtagtaaaa 594  
 QY 583 tatcgacgcacaagacacagggcgacgctactcagtgaggagccatgctatgca 642  
 DB 595 tatagcgcacaaccgcaagcgacacgctactcagtgaggagccatgctatgca 654  
 QY 643 cagggtcattttccaaacgctttggtatcacacacagctgaagctgctattccttg 702  
 DB 655 cagggtcattttccaaacgctttggtacacgggcaacagctaaaggctctattccttg 714  
 QY 703 gcagtagtcttgcgcgccaaagccacaaaacattctcaatcgctgtgtggttccaatgca 762  
 DB 715 gcagtggtcttgcgcgccagcgagggccaatctcaaatcgctgtgtgtctgtagca 774  
 QY 763 aaagctaacgctttgcagcgacagccattggtggaatactgtagtttaatttgggtcga 822  
 DB 775 acatctagctgttggagcgatagccttgggtgcaggtactcgtctcagctacagggc 834  
 QY 823 ggcttgccttagtcttgcctcagatcctgtgatagggtgataataataacagatgc 882  
 DB 835 agtattgccttagtcaaggtctgtgtcactcagagtgataataatc---tagaccg 891  
 QY 883 gcctatgtacactagttaaaacgttagcagacagctataaagccaccccgaggggtgat 942  
 DB 892 gcttatcaccaataacccgggcactagaccaccaagtttcaagccac---caataatcag 948  
 QY 943 tctacgatatattttccattggttaatagtaataataataatagcagtcagcgctaaa 1002  
 DB 949 aaggcgggtccacttccattgtagtaa-----ctctatcaaacgtaaa 993  
 QY 1003 atcatcaatgctcggtgcgggttctcgggatacccgatgcggtcaatgtggcacagcttaa 1062  
 DB 994 atcatcaatgctcggtgcgggttcttaataaaacccgatgcggtcaatgtggcacagtagaa 1053  
 QY 1063 ttggtgaggaacgtgcta---atcgtaaaattacttttaagggtgaggtgagcaataat 1119  
 DB 1054 gcgggtggtgaggtgggtcaaggcgtagaataattcttttcagggtgtagt-----ataac 1107  
 QY 1120 agcaatagctgagaagaggttttgggcaatactttaactttaagggtgagtcagacagcc 1179  
 DB 1108 agtactgacgtataaataaggtttggaataactttaactttaagggtgagtcagacagcc 1167  
 QY 1180 aacgcataaccgcaagctaacatcggtgtggtgaacagatggcaa-----tgggtctgaaa 1233  
 DB 1168 aacgcataaccgcaataatcggtgtggtgaagagggctgataatagtggtctgaaa 1227  
 QY 1234 gttaaacttgtaagagctgactggtggtggtggtggtggtggtggtggtggtggtggt 1278  
 DB 1228 gttaaacttgtaagagctgactggtggtggtggtggtggtggtggtggtggtggtggt 1287







	QY	5632	aactgtccatcagtgacgagcacacacgcaggccacacaagccaaaaattctgacggcaca	5691
	Db	5614	aactgtccatcagtgacgagcacacacgcaggccacacaagccaaaaattctgacggcaca	5673
	QY	5692	gcaggtacaaccacacacacgcaggtgccacaggtacgcggttaaaggctttgctggacaacg	5751
	Db	5674	gcaggtacaaccacacacacacgcaggtgccacacgcaggtgaacccggtacgcggttaaaggctttgctggacaacg	5733
	QY	5752	gcggttgggtgcggttcctcggtgggtgcctcaggtgctgaacccgtatccaaaatgtggca	5811
	Db	5734	gcggttgggtgcggttcctcggtgggtgcctcaggtgctgaacccgtatccaaaatgtggca	5793
	QY	5812	gcaggtgaggtcagtgcccaccagccgatgcggtccaatggtgtagccagtgtacaaagcc	5871
	Db	5794	gcaggtgaggtcagtgcccaccagccgatgcggtccaatggtgtagccagtgtacaaagcc	5853
	QY	5872	accaagaagcatggccaacgcacccaatgagctgacctgacctgataccacaaacgaaaaat	5931
	Db	5854	accaagaagcatggccaacgcacccaatgagctgacctgacctgataccacaaacgaaaaat	5913
	QY	5932	aaagccaatgcaggtatttcacagcgatggcgatggcgctccatgccacaagcctacatt	5991
	Db	5914	aaagccaatgcaggtatttcacagcgatggcgatggcgctccatgccacaagcctacatt	5973
	QY	5992	cctggcagatccatggttacccggggtattgccaacccacacacgcgtcaaggtgcggtggca	6051
	Db	5974	cctggcagatccatggttacccggggtattgccaacccacacacgcgtcaaggtgcggtggca	6033
	QY	6052	gtggagcgtgcgaagctgcggataatggtccaatgggtatttaaaatcaatggttcagcc	6111
	Db	6034	gtggagcgtgcgaagctgcggataatggtccaatgggtatttaaaatcaatggttcagcc	6093
	QY	6112	gatacccaagccatgtagggggcgagctgggtgcaggttttcacitt	6159
	Db	6094	gatacccaagccatgtagggggcgagctgggtgcaggttttcacitt	6141
	RESULT	3		
	ID	AAF59102		
	TD	AAF59102 standard; DNA; 6972 BP.		
	XX	AC		
	XX	AAF59102;		
	DT	24-APR-2001 (first entry)		
	XX	M. catarrhalis strain 4223 genomic 200kba gene SEQ ID NO:5.		
	DE			
	XX	Moraxella catarrhalis strain 4223; major outer membrane protein;		
	KW	200kDa outer membrane protein; antibacterial; immunogenic; infection;		
	KW	otitis media; detection; ds.		
	KX			
	OS	Moraxella catarrhalis.		
	XX			
	PN	WO200107619-A1.		
	XX			
	PD	01-FEB-2001.		
	XX			
	PF	26-JUL-2000; 2000WO-CA00870.		
	XX			
	PR	27-JUL-1999; 99US-0361619.		
	XX	(CONN-) CONNAUGHT LAB LTD.		
	PA	Loosmore SM, Sasaki K, Yang Y, Klein MH;		
	XX			
	PI			
	XX	WPI: 2001-159722/16.		
	DR	P-PSDB; AAB69134.		
	XX	New nucleic acid encoding Moraxella catarrhalis outer membrane protein,		
	PT	useful in protective vaccines and for diagnosis		
	XX			
	PS	Claim 1; Fig 3A-W; 247pp; English.		

The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of *Moraxella catarrhalis*. The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II)' and its truncated versions, are used as immunogenic compositions and vaccines to protect against *M. catarrhalis* infections, particularly otitis media in humans. (II) is also used as antigen in immunoassays for detecting specific antibodies (Ab), and to generate Ab. (I) are used for recombinant production of (II) and its fragments are used as probes for identifying/cloning 200 kDa protein genes from other strains, and for diagnostic detection of *M. catarrhalis*. (I) makes possible production of large amount of recombinant immunogens. Expression of truncated versions of (II) reduces toxicity of the protein towards the *Escherichia coli* host. The present sequence represents the *M. catarrhalis* strain 4223 genomic 200kDa gene, which is given in the exemplification of the present invention.

Sequence 6972 BP; 2265 A; 1555 C; 1532 G; 1620 T; 0 other;

Query Match	66.6%	Score 4101.6	DB 22	Length 6972	
Best Local Similarity	81.3%	Pred. No. 0			
Matches 5063	Conservative	0	Mismatches 1009	Indels 156	Gaps 20
QY 1	atgaatcaacatctataaagtcattctttaacaaagccacagggcacatttatggccgtggcg	60			
DB 542	atgaatcaacatctataaagtcattctttaacaaagccacagggcacatttatggcagtgcca	601			
QY 61	gaatagcacaaatccacagctacaggggggggtgctgtctacagggcaagttggcagt	120			
DB 602	gagtaagccaaatccacagctacaggggggggtgctgtctacagggcaagttggcagt	661			
QY 121	gtacgcaactctaaagctttggccgttatctccgcgcctgcctctctgtatcggtgcgcag	180			
DB 662	gtatgcactctgagctttggccgttatctccgcgcctgcctctctgtatcggtgcgcagcg	721			
QY 181	ctcaatggcagctcttatgtctcaacaaatctactacca---agatcgcaaatgtgtcaaac	237			
DB 722	ctcagtgggcgtgcttatgtctcaaaaaagataccaaacatatcgcaattgtgtgaacca	781			
QY 238	aacaagataaacacacgcctgaaagcgcatgccctagcgacaggtgaagcatccattgct	297			
DB 782	aaccagccaagacgcctc-----adggactgcgaaggcggagcggtgatcagagccattgct	835			
QY 298	tctggtagctttcttaaggcacaagcctctcaagcttatgtctatcggttagtgtcaaacca	357			
DB 836	atttgtgaaatgctcaacgcacagggcggtcgaagccatcgccatcggtagtagtaataaa	895			
QY 358	gatcctaataatgtagtaataatgtaagttaggtgtccacgcgaaggtaaacagatccatc	417			
DB 896	actgtcaatggaagcagtttggtaagataggttacogtatctcaggttcaagagtcctatc	955			
QY 418	gccatcggtgtgtgtatgtgctggaggtgtagctcgtgattgccatcggtagtgatgcac	477			
DB 956	gccatcggtgtgtgtataaaggctagtggtgagtcgtgatggccatcggttagtgatgcac	1015			
QY 478	ttatatcttgcctaagaatcttgcattctgaagaa-----tgaatttccacaaa	522			
DB 1016	ttcaatttgcctgtatcagcatgtgaacctcaacatccgaaggttactctgattaaacgat	1075			
QY 523	cttattcatggccatgaaataattaaaaaaaatacaaacctcaaccgatggtaaaatcaaaa	582			
DB 1076	cttattcaacggcatgcagttataaagaataacgaagctcaagaagtaatgtatgataaaa	1135			
QY 583	tatcgacgcaagaagcacaagggcagccagctactgcagtgagggaccatgtcatatgca	642			
DB 1136	tatagacgcaaacccgcgaagcggacagccagctactgcagtgagggccatgtcatatgca	1195			
QY 643	cagggtcatttttccaacgccttttggtaatacgaacacagctgaagctgcctattcccttg	702			
DB 1196	cagggtcatttttccaacgccttttggtaacagcgggaacagctaaaagtgccattcccttg	1255			
QY 703	gcagtaggtgtcttgcgcgcccaagcacaacaaatcttcaatcgtgtgtgttccaatgca	762			

XX  
PS  
Claim 1: Fig 3A-W; 247pp; English.

[illegible]

QY	2884	gcctcaaacacccctttacccgttaaaaggttagatgaaatgataaaggtgagacacacac	2943	QY	3961	aaagtgtctatgatgtcaatgtgatatacaacattgaagttaaagataaaaaactt	4020
Db	3398	gcctcaaacacccctttacccgttaaaaggttagatgaaatgataaaggtgagacacacac	3457	Db	4475	aaagtgtctatgagtcaatgtgatatacaacattgaagttaaagataaaaaactt	4534
QY	2944	gcataccagtggttaaaagatgacacaaagtgttaaaagtcacacaccccttaaaactcaaaaggt	3003	QY	4021	ggcgtaaaaaacacacacattgaccggtactggtgacaggtgctataataatttgcctcaagc	4080
Db	3458	gcataccagtggttcaaaagacgcaataatacaagtcaaacacccctcaacactcaaaaggt	3517	Db	4535	ggcgtaaaaaacacacacattgaccggtactggtgacaggtgctataataatttgcctcaagc	4594
QY	3004	aaaaacggtcttgatattaaaacccgacaaagatggtacaggttacccttggcattaaacac	3063	QY	4081	aatacagctactggcgatcgcttgcagggccagtgatctgcttgcctcatctataaacac	4140
Db	3518	gaaaacggtcttaattataaaacccgacaaatggtacaggttacccttggcattaaacac	3577	Db	4595	aatacagctactggcgatcgcttgcagggccagtgatctgcttgcctcatctataaacac	4654
QY	3064	caaaagcgtctaaagccgacgacacacacacacacacacacacacacacacacacacac	3123	QY	4141	ttatctggcgacatccaaactgcacaaaggggcgaagccaaagccaaactccagcaggtat	4200
Db	3578	acaagcgtctaaagccgacgacacacacacacacacacacacacacacacacacacac	3634	Db	4655	ttatctggcgacatccaaactgcacaaaggggcgaagccaaagccaaactccagcaggtat	4714
QY	3124	aacacccgtactgaacgac	3182	QY	4201	gtggatgctgataggcaataaggtcatctatgacagtcaccgatacaacagttactatcaagcc	4260
Db	3635	aacccacgtgtagcgac	3694	Db	4715	gtggatgctgataggcaataaaggctcatctatgacagtcaccgatacaacagttactatcaagcc	4774
QY	3183	--taataatggtgttagtgctggtgacatggtgacacacacacacacacacacacacacac	3240	QY	4261	aaaaatgatggcacagttgataaaacacacacacacacacacacacacacacacacacacac	4320
Db	3695	aataataatggtgttagtgctggtgacatggtgacacacacacacacacacacacacacac	3754	Db	4775	aaaaatgatggcacagttgataaaacacacacacacacacacacacacacacacacacacac	4834
QY	3241	attggctttactgggactaaatgggtcacttgataaaagcgaacacacacacacacacacacac	3300	QY	4321	gcccac	4380
Db	3755	attggctttactgggactaaatgggtcacttgataaaagcgaacacacacacacacacacacac	3814	Db	4835	gcccac	4894
QY	3301	ggcattacacgaggtgtgaaagattacacacacacacacacacacacacacacacacacacac	3360	QY	4381	caagttaaatgatgcacataaaagcgaacacacacacacacacacacacacacacacacacacac	4440
Db	3815	ggcattacacgaggtgtgaaagattacacacacacacacacacacacacacacacacacacac	3874	Db	4895	caagttaaatgatgcacataaaagcgaacacacacacacacacacacacacacacacacacacac	4954
QY	3361	agccatgagtgtgacagggcggaagatttatgatttaaaacacacacacacacacacacacacac	3420	QY	4441	cttgaaaaagcgcgtcttctgatacaaaaacacacacacacacacacacacacacacacacacacac	4500
Db	3875	agccatgagtgtgacagggcggaagatttatgatttaaaacacacacacacacacacacacacac	3934	Db	4955	cttgaaaaagcgcgtcttctgatacaaaaacacacacacacacacacacacacacacacacacacac	5014
QY	3421	atcagcagttactgcaaaaac	3480	QY	4501	aatgcgcgttggcccaaac	4560
Db	3935	atcagcagttactgcaaaaac	3994	Db	5015	aatgcgcgttggcccaaac	5074
QY	3481	caaggttaataactttacggttagtaac	3540	QY	4561	ctggcgagagactttgac	4620
Db	3995	caaggttaataactttacggttagtaac	4054	Db	5075	ctggcgagagactttgac	5134
QY	3541	gatgtcatcactttgcaggtgaaacgcgattacacacacacacacacacacacacacacacacacacac	3600	QY	4621	aacatcggtgtgtagcaggtactggttgcac	4680
Db	4055	gatgtcatcactttgcaggtgaaacgcgattacacacacacacacacacacacacacacacacacacac	4114	Db	5135	aacatcggtgtgtagcaggtactggttgcac	5194
QY	3601	cgtgtggcattgacaaac	3660	QY	4681	aatcttaacagcgttaagtgcaggtgac	4740
Db	4115	cgtgtggcattgacaaac	4174	Db	5195	aatcttaacagcgttaagtgcaggtgac	5254
QY	3661	aatggcaaaagcattgtcattac	3720	QY	4741	gacgcaaacggtcaagcccaaac	4800
Db	4175	aatggcaaaagcattgtcattac	4234	Db	5255	gactcaagcgttcaagcccaaac	5314
QY	3721	aacactctagcctaattgatacgaatgataaggttagcgtacacacacacacacacacacacacacacacac	3780	QY	4801	ggtggcaggttcacatgaatgtggcgaagcgaac	4860
Db	4235	aacactctagcctaattgatacgaatgataaggttagcgtacacacacacacacacacacacacacacacac	4294	Db	5315	ggtggcaggttcacatgaatgtggcgaagcgaac	5374
QY	3781	ataatcaaaagcgaagac	3840	QY	4861	caacagttaaacaggaagtcac	4911
Db	4295	ataatcaaaagcgaagac	4354	Db	5375	caacagttaaacaggaagtcac	5434
QY	3841	tttaacttgcgaagcaatgggtgaagcgttgaactttgctccacttatgacacacacacacacacacacacac	3900	QY	4912	gacggcaatcaggttaaacattgac	4971
Db	4355	tttaacttgcgaagcaatgggtgaagcgttgaactttgctccacttatgacacacacacacacacacacacac	4414	Db	5435	gacggcaatcaggttaaacattgac	5494
QY	3901	tttgcaatggcaatac	3960	QY	4972	aaccgactctcatcaaaagcaggtgacttgcgttaaaaggttaaaaggttaaaaggttaaaaggttaaaaggttaaaaggt	5031
Db	4415	tttgcaatggcaatac	4474	Db	5495	aaccgactctcatcaaaagcaggtgacttgcgttaaaaggttaaaaggttaaaaggttaaaaggttaaaaggttaaaaggt	5554
QY	4055	tttgcaatggcaatac	4119	QY	5032	aaacttgcacactggtggtgtacaaagtggcgctggataaaagcgcgaacgcgttaacgcgcgt	5091



1767 agttaaacttgctaaacttttaaacaactcttactgaggtggaatacaactacaattaaatgac 1826  
1278 taccacaataatcacggttagtaataccaacaacaacaaacgcgcgagctacaagcgggtgg 1337  
1827 cacaccacagcttaaggtaggttagtagtagtactacagctgaattattgagtgatag 1886  
1338 tttagacctttagccca-----ataacagggtacaataaaacagataaaaaccgtctacagcat 1391  
1887 tttaacctttaccagcccaatacagcgagctcaagcagcaagcaaaaccgtctcatggcgt 1946  
1392 tgaatgattgaagtttactaatgatagtaaagtagtagcaactaaaaggtactactcgtat 1451  
1947 taatgggggtgaagtttactaatatgcaagaacaacagcagcaaatcgccactactcgtat 2006  
1452 taccaaaaagaataatgggtttgtcgtggtactaaatgtagggagttgtagaaagcaaacctta 1511  
2007 taccagagataaaattggcttgcgcgagttggtagttgagttgtagtaaaaaacaacacata 2066  
1512 tccttgacaacgaagctaaagttggcaacagcaccctaaacagtggtgagcttgaactgt 1571  
2067 ttgggataaaaaaacacttaagttgggtgagttgtcaattaccatagacaatggcattga 2126  
1572 taatacacaccactggtaataaaaaaatccaaacgcggtgctaatggcattaaattgccac 1631  
2127 tgcaggtaataaaaagatcagtaactcttgccaaagtagtcaggtgctaacgtagcggttac 2186  
1632 agtcgtataaatgttgcaaataccctcagcaacagtcggcactgctgtattaccgaaga 1691  
2187 catcgacagctcaagcgcgaagcctactttaaacgcaggcgtggtcagctgclac 2246  
1692 gaaaattgggtttgtcgtgtactaatgatggagttgatgaacaagcaccatattggataa 1751  
2247 acctacigaaatatcagttgatgctaagagtggaatgtttaccgcccacacttaacaa- 2305  
1752 aqaacgacttaaaatgggtcggttgaaattaccacagatagtggtatttaatgctggttaa 1811  
2306 -----ttggcgtgaaacaccaccagcttaacagtgatgagcacttagtgaataatttag 2357  
1812 tcacaagattaccggacttactaatggtatagcaataacgtagcggttaccatcaaca 1871  
2358 tgttaagggtaggtgtagcaaatagcttagttacgcgcgaacattggcgaagctatct 2417  
1872 gctcaaaagcccaagcctactttaaacgcagcgatggtcagctacagtatantagtaataa 1931  
2418 aatatgaatcnaatcgaaacgctgacagtgctctacaagctttaccgtttaagaagaaga 2477  
1932 cgggagcttagttgatagtaggcaatattaccaccccaacttaaacatlagcgtgaa 1991  
2478 cgatgatgag-----ccaagcgtataccggtgac 2507  
1992 aacctaaagcttaacgataatgccaacagctgggtaataataatttagtgttagtaatgc 2051  
2508 taagatagcgaataaaatgcccgcgcagtcagcatcttaaacctcaaaaggtataaacgg 2567  
2052 tcatgaatacaatagcttagttaccgcgaagatttggcagactatcaataaagtcacaa 2111  
2568 tctaaagcgtgtctaccaaaaagatggtacggttacctttgggtcgtgccaagatagcgg 2627  
2112 tgaacgcgtgacagtgctctaccagctttaaagtccaaaacgggtgataatgacaacaa 2171  
2628 tctgaccattggaagaagcacctaaacaacagctggtctgactgttaagataccacaga 2687  
2172 cgccatcacccgtgggtgaagaatacaaacgggaagccttcaacaccttaaacctcaag 2231  
2688 acaaatccaagtcggtgctaattggcaataaatttactaatgtgaatggtaglaatccagg 2747  
2232 tgaacacgggtgtaatttacgaccaatagaccacaggtacagttacctttggcattga 2291  
2748 tactggcattgcaataaccgctgcgatta---ccagagataaaattgcttctgtggttc 2804  
2292 ccaagtaatggtctccacacgcgctaagctgaccgtgggttagcatacaaatgglaatcg 2351



Db 2805 tgatggtcagttgatcac-----aaacaaaccttatcttgatcaagacaaagcta--ca 2855  
QY 2352 attggttattgagcaagtccttagctgacggttaacagacacaaacacatcatataag 2411  
Db 2856 agttgcaatgttaagattaccaacactggcattaacgcaggtggttaagccatcacag 2915  
QY 2412 attgtcccaaacactccttagcattgcccagtcacaagtgccgccaacatagcattgggcaa 2471  
Db 2916 gctgtcccaaacactccttagcattgcccagtcacaagtagccgccaacatagaactgggcaa 2975  
QY 2472 tacaatcgaagaaaaagacaaaatcccaacgtgccagcattgattgattgcttaaatcagg 2531  
Db 2976 tacaatcccaagacaaagacaaaatcccaacgtgccagcattgattgattgattgattgatt 3035  
QY 2532 ctttaacctaaaaataatgtgcaaaagacaaaagactttgtctccacttatgacactgttga 2591  
Db 3036 ctttaacctaaaaataatgaaacccccattgactttgtctccacttatgacactgttga 3095  
QY 2592 ctttatcgatggcaatgccacacccgccacagtaacttatgatgaagccaatcaaacccag 2651  
Db 3096 ctttggccaatggcaatggcaccacccgccacagtaacccatgatacccgctaaacaaaaccag 3155  
QY 2652 taaagtggcgatgattgtaattggtatgagaaaaacattgaaactgacagcgataatgg 2711  
Db 3156 taaagtggatgattgtaattggtatgatacaaacatttcataacaggccaactgatga 3215  
QY 2712 caagaacaaacttggcgctcaaaacccatcaaaactgacgaacaaagtactaatgttaatgc 2771  
Db 3216 caataaaaaaacttggcgctcaaaacccacaaactgaacaaacaaagtgtcaatggttaatc 3275  
QY 2772 aactaacat-----tagtaccgacgatgacatgacctgacctgtttaaagccagtgatat 2822  
Db 3276 agcaactaaactttaattgtaactctagtgtatgaagtgccttgtttaacgccaagaacat 3335  
QY 2823 cgcggcaacttaaaacccctagcgagaaattccacacacaaaggcacagcaaacac 2882  
Db 3336 cgcgaaaaacttaaaacccctagcgaaggaaattccacacacaaaggcacagcaaacac 3395  
QY 2883 cgccttacaacactttaccgctaaaaaggtagatgaaaatgataaaggctgatgacaccaa 2942  
Db 3396 cgccttacaacactttaccgctaaaaaggttagatgaaaataataatgctgatgacgcaa 3455  
QY 2943 cgcattcacccgtgggttaagatggcacaaggtgtaagtcacaacaccttaaacctcaaaag 3002  
Db 3456 cgcattcacccgtgggttaaaagacgcacaaataatcgaagtcaacacccctaaacacacaaag 3515  
QY 3003 taaaaacggtcttgattataaaacccgacaaaagatggtacggttacctttggcattaacac 3062  
Db 3516 tgaacacggtcttaattataaaacccgacaaaataatggttacggttacctttggcattaacac 3575  
QY 3063 ccaaaagcgtcttaagccggtgacagcaccactctaaacaaatggcttgtctattaa 3122  
Db 3576 cacaagcgtcttaagccggtgacaaagcacc---ctaaacgacggtggcttgcatttaa 3632  
QY 3123 aaacacccgtagtaagacaaataatcaagtcggtgctgattgagcgtgaagtttccatggt 3182  
Db 3633 aaacccacgttagcgaacaaataatcaagtcggtgctgattgagcgtgaagtttccaaaggt 3692  
QY 3183 ---taataatggttctgattgctggtcattgattggaacaaactcgattaccagagatga 3239  
Db 3693 taataataatggttctgattgctggtcattgattggaacaaactcgattaccagagatga 3752  
QY 3240 aattgcttactggactaatggctcactttgataaagcaaacccacacctaagcaaga 3299  
Db 3753 aattgcttactggactaatggctcactttgataaagcaaacccacacctaagcaaga 3812  
QY 3300 cggcattaaacgaggttggtaaaagattaccaacattcaatcagggtgagattgccaaaaa 3359  
Db 3813 cggcattaaacgaggttggtaaaagattaccaacattcaatcagggtgagattgccaaaaa 3872  
QY 3360 cagccatgattgctgacagggcggaagatttatgatttaaaacccgaacttgaataaa 3419  
Db 3873 cagccatgattgctgacagggcggaagatttatgatttaaaacccgaacttgaataaa 3932

QY 3420 aatcagcagtactgcgcaaaacagcacaaaactcattacacgaattctcagtagcagatga 3479  
Db 3933 aatcagcagtactgcgcaaaacagcacaaaactcattacacgaattctcagtagcagatga 3992  
QY 3480 acaaggtataaactttacggttagtaacccttactcagttatgacacccctcaaaagacctc 3539  
Db 3993 acaaggtataaactttacggttagtaacccttactcagttatgacacccctcaaaagacctc 4052  
QY 3540 tgaatgcatacactttgcaggtgaaaaacgcattaccaccaaaggtataaaggtgtggt 3599  
Db 4053 tgaatgcatacactttgcaggtgaaaaacgcattaccaccaaaggtataaaggtgtggt 4112  
QY 3600 gcgtgtgggcatgaccacaaacccaaagccttaaccacgcctaaagctgacctggtggtataaa 3659  
Db 4113 gcgtgtgggcatgaccacaaacccaaagccttaaccacgcctaaagctgacctggtggtataaa 4172  
QY 3660 taatggcaaaagcattgtcttaataacagccaaaatggccaataatccatcacagacataag 3719  
Db 4173 taatggcaaaagcattgtcttaataacagccaaaatggccaataatccatcacagacataag 4232  
QY 3720 caacactctagctaaattgtaccatgataaaggtagcgtacgcacccacagaaacagggcaa 3779  
Db 4233 caacactctagctaaattgtaccatgataaaggtagcgtacgcacccacagaaacagggcaa 4292  
QY 3780 tataatcaagaagacagacaaaaccccgccgcagcattgtgtgattgcttaagcgcag 3839  
Db 4293 tataatcaagaagacagacaaaaccccgccgcagcattgtgtgattgcttaagcgcag 4352  
QY 3840 ctttaacttgcagggcaatggtgaagcgttgactttgactttgctccacttatgacacgcgtcaa 3899  
Db 4353 ctttaacttgcagggcaatggtgaagcgttgactttgactttgctccacttatgacacgcgtcaa 4412  
QY 3900 ctttgccaatggcaatcaccaccccgctaaaggtgacctatgatgacacaaacaaacag 3959  
Db 4413 ctttgccaatggcaatcaccaccccgctaaaggtgacctatgatgacacaaacaaacag 4472  
QY 3960 taaagtgttctatgattgctcaatgtggatgatacaacacctgaagttaaagataaaaaact 4019  
Db 4473 taaagtgttctatgattgctcaatgtggatgatacaacacctgaagttaaagataaaaaact 4532  
QY 4020 tggcgtaaaaaacacacacattgaccagttactggtgacacaggtgcttaataaaattgcccctaag 4079  
Db 4533 tggcgtaaaaaacacacacattgaccagttactggtgacacaggtgcttaataaaattgcccctaag 4592  
QY 4080 caatcaagcttactggcgatgcgcttgcagggccagtgatgattgctcattctaaacac 4139  
Db 4593 caatcaagcttactggcgatgcgcttgcagggccagtgatgattgctcattctaaacac 4652  
QY 4140 cttatctggcgacatccaaactgccaaaggggcgaagccaaagcgaacaaactcagcagcta 4199  
Db 4653 cttatctggcgacatccaaactgccaaaggggcgaagccaaagcgaacaaactcagcagcta 4712  
QY 4200 tggcgatgctgattggaataaaggtcattctatgacagttaccgataacaaagtactatcaagc 4259  
Db 4713 tggcgatgctgattggaataaaggtcattctatgacagttaccgataacaaagtactatcaagc 4772  
QY 4260 caaaaaatgattggcacagttgataaaaaaccaaagagttgccaagaacaaactggtgcgcca 4319  
Db 4773 caaaaaatgattggcacagttgataaaaaaccaaagagttgccaagaacaaactggtgcgcca 4832  
QY 4320 agcccaaacccccagatggcacattggctcaaatgaaatgtaaatcagtcatttaacaaga 4379  
Db 4833 agcccaaacccccagatggcacattggctcaaatgaaatgtaaatcagtcatttaacaaga 4892  
QY 4380 acaagtaaaatgattgccaataaaaaagcaaggtcattcaatgaagacaaacgcttggtaaagg 4439  
Db 4893 acaagtaaaatgattgccaataaaaaagcaaggtcattcaatgaagacaaacgcttggtaaagg 4952  
QY 4440 acttgaaaagcgccttctgataaaaaacccaaacccgacgagtaactggtggtgattt 4499  
Db 4953 acttgaaaagcgccttctgataaaaaacccaaacccgacgagtaactggtggtgattt 5012



Thu Sep 13 14:19:05 2001

QY	4500	aaatgcgctgtgcccacacacgcgtgaccttgcaggggtacagggacacacgcgttaaaaa	4559		Db	6093	tgctttggtgtgggcaataaacatcacgtgacggaaagtaactcggttgcttaggttc	6152
Db	5013	aaatgcgctgtgcccacacacgcgtgaccttgcaggggtacagggacacacgcgttaaaaa	5072		QY	5631	aaactctgccatcagtcagggcacacacgcagggacacacgcgttaaaaaatctgacggcac	5690
QY	4560	actgggagagacttgcacatcaaatgggggcaaacagacacacataaagctaacacgataa	4619		Db	6153	aaactctgccatcagtcagggcacacacgcagggacacacgcgttaaaaaatctgacggcac	6212
Db	5073	actgggagagacttgcacatcaaatgggggcaaacagacacacataaagctaacacgataa	5132		QY	5691	agcaggtacacacacacacacaggtgcccacaggtacggttaaaaggcttgcgtggacacac	5750
QY	4620	taacatcggtgtgtagcaggtactgagcttgcacttgcacttgcacttgcacttgcacttgc	4679		Db	6213	agcaggtacacacacacacacaggtgcccacaggtacggttaaaaggcttgcgtggacacac	6272
Db	5133	taacatcggtgtgtagcaggtactgagcttgcacttgcacttgcacttgcacttgcacttgc	5192		QY	5751	ggcgtgtggtggt	5810
QY	4680	caatctaacagcgttaagtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgag	4739		Db	6273	ggcgtgtggtggt	6332
Db	5193	caatctaacagcgttaagtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgag	5252		QY	5811	agcaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggt	5870
QY	4740	agcgcacacggtgacgac	4799		Db	6333	agcaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggt	6392
Db	5253	agcgcacacggtgacgac	5312		QY	5871	cacccaaagcattgccaacgcaac	5930
QY	4800	gggtggcaggtgacgac	4859		Db	6393	cacccaaagcattgccaacgcaac	6452
Db	5313	gggtggcaggtgacgac	5372		QY	5931	taaaagcgaatgaggttgcac	5990
QY	4860	acaacaggttaaacgaggtgacgacacacacacacacacacacacacacacacacacacac	4910		Db	6453	taagggcgaatgaggttgcac	6512
Db	5373	acaacaggttaaacgaggtgacgacacacacacacacacacacacacacacacacacacac	5432		QY	5991	tcttggcagatccatggttaccgggggtattgcccacacacacacacacacacacacacacac	6050
QY	4911	tgacggcgaatgaggttaaacatgcccacacacacacacacacacacacacacacacacac	4970		Db	6513	tcttggcagatccatggttaccgggggtattgcccacacacacacacacacacacacacacac	6572
Db	5433	tgacggcgaatgaggttaaacatgcccacacacacacacacacacacacacacacacacac	5492		QY	6051	agtgggacgtgcaagctgctcggaataatggttcaatgggttcaatgggttcaatgggttcaat	6110
QY	4971	taacgcacgtgac	5030		Db	6573	agtgggacgtgcaagctgctcggaataatggttcaatgggttcaatgggttcaatgggttcaat	6632
Db	5493	taacgcacgtgac	5552		QY	6111	cgatacccaaggccatgtagggcgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggt	6159
QY	5031	aaaactgtgac	5090		Db	6633	cgatacccaaggccatgtagggcgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggt	6681
Db	5553	aaaactgtgac	5612		RESULT	5		
QY	5091	tttaagcaatgttgggttcaaaac	5150		AAF59100	standard; DNA; 6973 BP.		
Db	5613	tttaagcaatgttgggttcaaaac	5672		XX	AAF59100		
QY	5151	ttataacgcgcaggtgac	5210		AC			
Db	5673	ttataacgcgcaggtgac	5732		XX	24-APR-2001 (first entry)		
QY	5211	ataaatgacacaggtatccgttcttccatgtcaacaggtgacacacacacacacacacacac	5270		DT			
Db	5733	ataaatgacacaggtatccgttcttccatgtcaacaggtgacacacacacacacacacacac	5792		XX	M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa gene SEQ ID NO:1.		
QY	5271	acaagggcgaacggtgac	5330		DE			
Db	5793	acaagggcgaacggtgac	5852		XX	Moraxella catarrhalis strain 4223; major outer membrane protein;		
QY	5331	ccagggcgaacggtgac	5390		KW	200kDa outer membrane protein; antibacterial; immunogenic; infection;		
Db	5853	ccagggcgaacggtgac	5912		KW	otitis media; detection; ds.		
QY	5391	ccaatccatgcac	5450		XX	Moraxella catarrhalis.		
Db	5913	ccaatccatgcac	5972		OS			
QY	5451	tacagggcgaacggtgac	5510		PN	WO200107619-A1.		
Db	5973	tacagggcgaacggtgac	6032		PD	01-FEB-2001.		
QY	5511	ggcgtgataac	5570		PF	26-JUL-2000; 2000WO-CA00870.		
Db	6033	ggcgtgataac	6092		XX	27-JUL-1999; 99US-0361619.		
QY	5571	tgctttggtgtgggcaataac	5630		PR	(CONN-) CONNAUGHT LAB LTD.		
Db					PA			
QY					XX	Loosmore SM, Sasaki K, Yang Y, Klein MH;		
Db					PI			
QY					XX	WPI; 2001-159722/16.		
Db					DR	P-PSDB; AAB69133.		
QY					XX	New nucleic acid encoding Moraxella catarrhalis outer membrane protein,		
Db					PT	useful in protective vaccines and for diagnosis		
QY					PT			

XX

Example 3; Fig 2A-W; 247pp; English.

The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of *Moraxella catarrhalis*. The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II), and its truncated versions, are used as immunogenic compositions and vaccines to protect against *M. catarrhalis* infections, particularly otitis media in humans. (II) is also used as antigen in immunoassays for detecting specific antibodies (Ab), and to generate Ab. (I) are used for recombinant production of (II) and its fragments are used as probes for identifying/cloning 200 kDa protein genes from other strains, and for diagnostic detection of *M. catarrhalis*. (I) makes possible production of large amount of recombinant immunogens. Expression of truncated versions of (II) reduces toxicity of the protein towards the *Escherichia coli* host. The present sequence represents the *M. catarrhalis* strain 4223 lambdaDEMBL3 clone 200kDa gene, which is used in the exemplification of the present invention.

SQ Sequence 6973 BP; 2265 A; 1555 C; 1533 G; 1620 T; 0 other;

Query Match 66.4%; Score 4090.6; DB 22; Length 6973;  
Best Local Similarity 81.3%; Pred. NO. 0;  
Matches 5063; Conservative 0; Mismatches 1009; Indels 157; Gaps 21;

Qy 1 atgaatcacatctataaagtcattctttaacaaagccacaggcacatttatggccqtqccq 60

Db 542 atgaatcacatctataaagtcatcttttaacaaagccacaggcacatttatggcagtgcca 601

Qy 61 gaatatgccaaatccccacagtac-qqqqqqqqqtagcttqctacagqqcacaagtqqcaq 119

Db 602 gagtacgccaaatccacagcacaggggggggtagctgtctacagggcaattggcag 661

Qy 120 tqtacqcactctaaqcctttqcccqatattqccqccctcgcctgcctcctcgtqatcgcgtgcgcac 179

Db 662 tgtatgcactctgagctttgcccgctattgccgcgcctgcctctctctgtatcggtgcaac 721

Qy 180 gctcaatggcaqtgctttatgctcaacaaattactacca--agatcqaatttqqtcacaac 236

Db 722 gctcagtggcagtgcttatgctcaaaaaaaagataccacaacatatcgccaattgggtgaaca 781

Qy 237 aaacaagataaacacacgcctgaaaggcgatgccctagcgacacaggtgaagcatccattgc 296

Db 782 aaaccagccaagacgcctc-----aggcactgccaaagcggcagctgatcgagccattgc 835

Qy 297 ttttggtagtccttcttaaggcacaaaggctctcaagctattgctatcggtagtggtcaaac 356

Db 836 tattggtgaaatgctaacgcacagggcggtcaagccatcgccatggttagtagtaataa 895

Qy 357 agatcctaataatggtagtaatggtaatgtaggttcccacgcccacgaagtccat 416

Db 896 aactgtcaatgggaagcagtttggataagataagggtaccatgtctacgggtcaagaagtccat 955

Qy 417 cgccatcggtggtgatgtattggctgaggggtgatgcctcgattgccatcggtagtgatga 476

Db 956 cgcacatcggtggtgatgtaaaggcctagtgtgtagtcgcctgattgccatcggtagtgatga 101

Qy 477 cttatatattgcctaagaatcttgatctgaagaa-----tgaatttcacaa 521

Db 1016 cttaacatttgcttgatcagcatggaatactctaaacatccgaaaggtactctgtattaacga 107

Qy 522 acttattcatggccatgaatatataaaaaatacaaacctcaaccqatqqtaaaatcaa 581

Db 1076 tcttattaaaggcccatgcagttataaaagaatacgaagctcaaggataatcgatgtaaa 1133

Qy 582 atatcgacgcacaaagacacaaaggccacgctactcgaqtqqgaqccatgtcatatgc 641

Db 1136 atatagacgcacaccccaagcgacacgcagctactgcagtggagccatctcatatgc 1199

Qy 642 acagggtcattttccaacgcctttggtacatacacacagctqaagctgcctattcctt 701

Db 1196 acagggtcatttttccaacgccttttggtacacggggcaacagctaataagtgcctatttcctt 125

[illegible]









[illegible]



Qy 5552 cacacagcagggcacacacagccaaataatctgacggcacagcaggtacacacacacacagc 5711  
 |||||  
 Db 48024 cacacagcagggcacacacagccaaataatctgacggcacacagcaggtacacacacacagc 48083  
 |||||  
 Qy 5712 aggtgccacaggtacggttaaaagcttctgctggacaaacgagcgttgggtgctccgt 5771  
 |||||  
 Db 48084 aggtgccacaggtacggttaaaagcttctgctggacaaacgagcgttgggtgctccgt 48143  
 |||||  
 Qy 5772 ggggtccacaggtacggttaaaagcttccaaatgtggcagcaggtgaggtcagtgccac 5831  
 |||||  
 Db 48144 ggggtccacaggtacggttaaaagcttccaaatgtggcagcaggtgaggtcagtgccac 48203  
 |||||  
 Qy 5832 cagcagcagtgctgaatggtagcaggtgtacaaagcagcagcagcagcagcagcagcagc 5891  
 |||||  
 Db 48204 cagcagcagtgctgaatggtagcaggtgtacaaagcagcagcagcagcagcagcagcagc 48263  
 |||||  
 Qy 5892 aaccaatgagcttgaccatgctatcccaaaacgaaataaaagcagcagcagcagcagcagc 5951  
 |||||  
 Db 48264 aaccaatgagcttgaccatgctatcccaaaacgaaataaaagcagcagcagcagcagcagc 48323  
 |||||  
 Qy 5952 atcagcagtgctgagcgtccatgccacaaagcctacattctgctgagcagcagcagcagcagc 6011  
 |||||  
 Db 48324 atcagcagtgctgagcgtccatgccacaaagcctacattctgctgagcagcagcagcagcagc 48383  
 |||||  
 Qy 6012 cgggggtattgccaccacacacagcgtcaggtgcggtggcagtggtgagcagcagcagcagcagc 6071  
 |||||  
 Db 48384 cgggggtattgccaccacacacagcgtcaggtgcggtggcagtggtgagcagcagcagcagcagc 48443  
 |||||  
 Qy 6072 ggataatgggtcaatgggtatttaaaataaatggttcagcagcagcagcagcagcagcagcagc 6131  
 |||||  
 Db 48444 ggaataatgggtcaatgggtatttaaaataaatggttcagcagcagcagcagcagcagcagcagc 48503  
 |||||  
 Qy 6132 ggcgcagtggtgagcaggttttcaacttt 6159  
 |||||  
 Db 48504 ggcgcagtggtgagcaggttttcaacttt 48531  
 |||||

## RESULT 7

AAF59101

ID AAF59101 standard; DNA; 5979 BP.

AC AAF59101;

XX AAF59101;

DT 24-APR-2001 (first entry)

DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa coding sequence #2.

KW Moraxella catarrhalis strain 4223; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KW otitis media; detection; ds.

XX Moraxella catarrhalis.

OS Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

XX Moraxella catarrhalis.

xx

CC The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines, (II), and its truncated versions, are used as immunogenic compositions and vaccines to protect against M. catarrhalis infections, particularly otitis media in humans. (II) is also used as antigen in immunoassays for detecting specific antibodies (Ab), and to generate Ab. (I) are used for recombinant production of (II) and its fragments are used as probes for identifying/cloning 200 kDa protein genes from other strains, and for diagnostic detection of M. catarrhalis. (I) makes possible production of large amount of recombinant immunogens. Expression of truncated versions of (II) reduces toxicity of the protein towards the Escherichia coli host. The present sequence represents the M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa coding sequence, which is used in the exemplification of the present invention.

xx

SQ Sequence 5979 BP; 1974 A; 1360 C; 1355 G; 1290 T; 0 other;

Query Match 64.1%; Score 3947.8; DB 22; Length 5979;  
 Best Local Similarity 80.9%; Pred. No. 0;  
 Matches 4905; Conservative 0; Mismatches 1002; Indels 156; Gaps 20;

Qy

166 gtgatcggtgcgacgctcaatggcagtgcttatgtctcaacaaatattactaccag---atc 222

Db

1 gtgatcggtgcgacgctcagtgccagtgcttatgtctcaacaaatattactaccatc 60

Qy

223 gaaattggtcaacaaacagataaaacacacgctgaaggcgatgcctagcagcaggt 282

Db

61 gcaattggtgaacaaacagcagcagcagcagcagcagcagcagcagcagcagcagc 114

Qy

283 gaagcatcattgctttgtagtcttcttaaggacacagcgtctcaagcttattgtctatc 342

Db

115 gatcgagcattgctattggtgaaatgctaacgcacagcggtgcaagcagcagcagcagc 174

Qy

343 ggtatggtcaacacagatccttaataatggtatgtaattgtaattggttccacacgcaaa 402

Db

175 ggtatgataataaaactgtaattggaagcaggttggataagataggtagcagtgctacg 234

Qy

403 ggtacagagtcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 462

Db

235 ggtcaagagtcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 294

Qy

463 atcggtagtgatgacttatttgcctaaagaatcttgcctgaagaa----- 509

Db

295 atcggtagtgatgacttatttgcctaaagaatcttgcctgaagaa----- 354

Qy

510 --tgaattccaaacttatttgcctaaagaatcttgcctgaagaa----- 567

Db

355 actctgataacgacttatttgcctaaagaatcttgcctgaagaa----- 414

Qy

568 gatggtataaatcaaatatcgacgcacagacagcagcagcagcagcagcagcagcagcagc 627

Db

415 gataatgagtaaaataatcgacgcacagcagcagcagcagcagcagcagcagcagcagcagc 474

Qy

628 gccatgcatatgcacaggggtcattttccaaagcgtttggtacacacagcagcagcagcagc 687

Db

475 gccatgcatatgcacaggggtcattttccaaagcgtttggtacacacagcagcagcagcagc 534

Qy

688 gctgcctattccttgccagtgaggtcttgcgcacacagcagcagcagcagcagcagcagcagc 747

Db

535 agtgcctattccttgccagtgaggtcttgcgcacacagcagcagcagcagcagcagcagcagc 594

Qy

748 gttggttccaaatcaaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 807

Db

595 attggttctgatgcacacatcagcctgttggagcgatagcctctggtgaggtactcgt 654

Qy

808 gttaaattgggtcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 867

Db

655 gctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 714

Qy

868 aatacagatgcccagtcctatgtaccactaggtataaacagcagcagcagcagcagcagcagc 927

[illegible]

Qy 3049 ttggcattaacaccacccaaagcgggtcttaaacggcggcagacacacactcttaaacacaaat 3108  
Db 2857 ttggcattaacaccacacagcgggtcttaaacggcggcagacacacactcttaaacacaaat 2913  
Qy 3109 ggcgttctataaaaaacaccgcctagtaaacgacacaaatcccaagtcggctgctgagtgccgtg 3168  
Db 2914 ggcgttctataaaaaacccacactggtagcgaacaaatcccaagtcggctgctgagtgccgtg 2973  
Qy 3169 aagtttgcactgggt---taataaagggttggtagtgctgctgctgacacactgcg 3225  
Db 2974 aagtttgcactgggttaaataaagggttggtagtgctgctgctgacacactgcg 3033  
Qy 3226 attaccagagatgaattgggtttactgggactaatggctcaacttgataaagaacaccc 3285  
Db 3034 attaccagagatgaattgggtttactgggactaatggctcaacttgataaagaacaccc 3093  
Qy 3286 caccataagcaagacggcattaaacgacgtgtaaaagattaccaacattcaatcaggt 3345  
Db 3094 caccataagcaagacggcattaaacgacgtgtaaaagattaccaacattcaatcaggt 3153  
Qy 3346 gagattgccaacaaacagccatgactgtgacagggcggcagatttatgatttaaaacc 3405  
Db 3154 gagattgccaacaaacagccatgactgtgacagggcggcagatttatgatttaaaacc 3213  
Qy 3406 gaacttgaataataaatacagcagtagtctccaaaacagacaaaactattacaagaattc 3465  
Db 3214 gaacttgaataataaatacagcagtagtctccaaaacagacaaaactattacaagaattc 3273  
Qy 3466 tcaatgacagatgaacaaaggttaataaactttacggttagtaaaccttactccagttatgac 3525  
Db 3274 tcaatgacagatgaacaaaggttaataaactttacggttagtaaaccttactccagttatgac 3333  
Qy 3526 acctcaagacctctgactatcacctttgcaggtgaaaacggcatttaccacacaggtta 3585  
Db 3334 acctcaagacctctgactatcacctttgcaggtgaaaacggcatttaccacacaggtta 3393  
Qy 3586 aataaagggttggctgctgtggcattgacaaaaccaaaggccttaaccacgcctaaagctg 3645  
Db 3394 aataaagggttggctgctgtggcattgacaaaaccaaaggccttaaccacgcctaaagctg 3453  
Qy 3646 accgtgggttaataataatggcaaaagcattgtcattaaacagcacaataatgggtcacaataacc 3705  
Db 3454 accgtgggttaataataatggcaaaagcattgtcattaaacagcacaataatgggtcacaataacc 3513  
Qy 3706 atcacagactaaagcaacactctagctaatgttaccaatgataaaggtagcgtacgcacc 3765  
Db 3514 atcacagactaaagcaacactctagctaatgttaccaatgataaaggtagcgtacgcacc 3573  
Qy 3766 acagaacaggccaataataatcaaaagacgaagacaaaacccgtgcccgcagcatttgtat 3825  
Db 3574 acagaacaggccaataataatcaaaagacgaagacaaaacccgtgcccgcagcatttgtat 3633  
Qy 3826 gtgctaaagcagaccttaacttgaagcgaatggtgaagcgttgaacttgtctccact 3885  
Db 3634 gtgctaaagcagaccttaacttgaagcgaatggtgaagcgttgaacttgtctccact 3693  
Qy 3886 tatgacacgttcaactttgccaatggcgaataccacacacgcctaaaggtgacctatgac 3945  
Db 3694 tatgacacgttcaactttgccaatggcgaatgccacacacgcctaaaggtgacctatgac 3753  
Qy 3946 acaagcaaaacagtaaaagtggtctatgattcaatgtggatgatacaaacacattgaagtt 4005  
Db 3754 acaagcaaaacagtaaaagtggtctatgattcaatgtggatgatacaaacacattgaagtt 3813  
Qy 4006 aaaaataaaaacttggcgttaaaaacccacattgacaggttgcacaggtgctaat 4065  
Db 3814 aaaaataaaaacttggcgttaaaaacccacattgacaggttgcacaggtgctaat 3873  
Qy 4066 aaatttgccttaagcaatacgaactactggcgtgcttgaagggcaggtgatactgt 4125  
Db 3874 aaatttgccttaagcaatacgaactactggcgtgcttgaagggcaggtgatactgt 3933

Qy 4126 gctcatctaaacacacacttatctggcgacatccaaactgccaaggggcgaagccaagcgaac 4185  
Db 3934 gctcatctaaacacacacttatctggcgacatccaaactgccaaggggcgaagccaagcgaac 3993  
Qy 4186 aactcagcaggtctatgtgtagctgacgaataaggtctatctatgacagtagtaccgataac 4245  
Db 3994 aactcagcaggtctatgtgtagctgacgaataaggtctatctatgacagtagtaccgataac 4053  
Qy 4246 aagtactatcaagcgaacaaatgatggcacagttgataaaacccaaagaagttgtccaaagac 4305  
Db 4054 aagtactatcaagcgaacaaatgatggcacagttgataaaacccaaagaagttgtccaaagac 4113  
Qy 4306 aaactggctgcccgaagcccaacccacagatggcacattggctcacaatgaatgtcaaatca 4365  
Db 4114 aaactggctgcccgaagcccaacccacagatggcacattggctcacaatgaatgtcaaatca 4173  
Qy 4366 gtcatcaagaagaacaagtataatgatgccaataaaagcaggtcatcaatgaagaacac 4425  
Db 4174 gtcatcaagaagaacaagtataatgatgccaataaaagcaggtcatcaatgaagaacac 4233  
Qy 4426 gctttgttaaaggacttgaaaaagccgtctctgtataacaaacccaaacccgcagta 4485  
Db 4234 gctttgttaaaggacttgaaaaagccgtctctgtataacaaacccaaacccgcagta 4293  
Qy 4486 actgtgggtgatttaaatgcccgttccccaaacacgcgtgacctttgcaggggatacaggc 4545  
Db 4294 actgtgggtgatttaaatgcccgttccccaaacacgcgtgacctttgcaggggatacaggc 4353  
Qy 4546 acaacggctaaaaaactggcgagactttgacatcaaaaggtgggcaacacagacacacaaat 4605  
Db 4354 acaacggctaaaaaactggcgagactttgacatcaaaaggtgggcaacacagacacacaaat 4413  
Qy 4606 aagctaaacggataataacatcggtgtggtagcaggttactgtggttgcactgtccaaactt 4665  
Db 4414 aagctaaacggataataacatcggtgtggtagcaggttactgtggttgcactgtccaaactt 4473  
Qy 4666 gccaaagacctaaacacacttaacagcgttaataagcgtggcaccacaaatattgatgaaaa 4725  
Db 4474 gccaaagacctaaacacacttaacagcgttaataagcgtggcaccacaaatattgatgaaaa 4533  
Qy 4726 ggcattctcttggtagacgaacacggctcaagcgaagcacaacacccctgtctaaagtcgc 4785  
Db 4534 ggcattctcttggtagacgaacacggctcaagcgaagcacaacacccctgtctaaagtcgc 4593  
Qy 4786 aatggctgggaacctgggtggcgaaggtcatcagtaattgtgggcaaaagcacaagatacc 4845  
Db 4594 aatggctgggaacctgggtggcgaaggtcatcagtaattgtgggcaaaagcacaagatacc 4653  
Qy 4846 gacgtgccaatgtacaacagttaaacgaagtacgcaacttgttgggtct-----t 4896  
Db 4654 gacgtgccaatgtacaacagttaaacgaagtacgcaacttgttgggtcttggtaatgct 4713  
Qy 4897 ggaatgataacgcctgagcggcaatcaggtaaacattgcccacatcaaaaaagacacaaat 4956  
Db 4714 ggaatgataacgcctgagcggcaatcaggtaaacattgcccacatcaaaaaagacacaaat 4773  
Qy 4957 tcaggttcatcatctaaacgcactgtcatcaaaagcaggtcacgcgttacttggcgttaaggt 5016  
Db 4774 tcaggttcatcatctaaacgcactgtcatcaaaagcaggtcacgcgttacttggcgttaaggt 4833  
Qy 5017 aatacagatcacgaacaaacttgcactgttgggtgatacaagttgggcttgataaagcgcgc 5076  
Db 4834 aatacagatcacgaacaaacttgcactgttgggtgatacaagttgggcttgataaagcgcgc 4893  
Qy 5077 aacgtcaacgcgatttaagcaatgttgggtcaaaacccaaagaagttggcgaacaaa\* 5136  
Db 4894 aacgtcaacgcgatttaagcaatgttgggtcaaaacccaaagaagttggcgaacaaa 4953  
Qy 5137 gccctgctgcacacttaaacgcgcaggtcagacacactatgtgaccacacaccccgca 5196  
Db 4954 gccctgctgcacacttaaacgcgcaggtcagacacactatgtgaccacacaccccgca 5013  
Qy 5197 gaagccattgacagaataaataaaggtatccgcttcttccatgtcaacgatggcaat 5256

Thu Sep 13 14:19:05 2001

```

|||||
Db 5014 gaagccatgacagaataaataagaaaggtatccgctcttccatgtcaacgatggcaat 5073
QY 5257 caagagcctgtgtgtacaaagggcgaacgcttactcaagtgcctcaggcaagcactca 5316
Db 5074 caagagcctgtgtgtacaaagggcgaacgcttactcaagtgcctcaggcaagcactca 5133
QY 5317 gtgagcatagtttccagggcgaagcagatggtgaagcgcgcttgccatagcagacaa 5376
Db 5134 gtgagcatagtttccagggcgaagcagatggtgaagcgcgcttgccatagcagacaa 5193
QY 5377 accaaagcagggcaacaaatccatccatcgcgtgataacgcacaagccacggcgatcaa 5436
Db 5194 accaaagcagggcaacaaatccatccatcgcgtgataacgcacaagccacggcgatcaa 5253
QY 5437 tccatgccatcgtgtacaggaatgtgtgtgtagcaggtgaacactgtgtgcccacggcag 5496
Db 5254 tccatgccatcgtgtacaggaatgtgtgtgtagcaggtgaacactgtgtgcccacggcag 5313
QY 5497 ccaagcactgttaaggtgataacagttacagttggtgtggttaataacacacagttaccgat 5556
Db 5314 ccaagcactgttaaggtgataacagttacagttggtgtggttaataacacacagttaccgat 5373
QY 5557 gccactaaacccgatgtcttgggtgtgggaataacataccgtgacaaagtaactcg 5616
Db 5374 gccactaaacccgatgtcttgggtgtgggaataacataccgtgacaaagtaactcg 5433
QY 5617 gtgtccttaggttcaaacctctgcaatcagtgtaggcaggtgaacactgtgtgcccacggc 5676
Db 5434 gtgtccttaggttcaaacctctgcaatcagtgtaggcaggtgaacactgtgtgcccacggc 5493
QY 5677 aaatctgaagcagcagcaggttacacccacacagcagcagcagcagcagcagcagcagc 5736
Db 5494 aaatctgaagcagcagcaggttacacccacacagcagcagcagcagcagcagcagcagc 5553
QY 5737 ttgtcgtgcaaacagcagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5796
Db 5554 ttgtcgtgcaaacagcagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5613
QY 5797 atccaaatgtgtgagcagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5856
Db 5614 atccaaatgtgtgagcagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5673
QY 5857 cagttgtacaaagccacccaaagcattgcacacgcaacacaaatgagcttgaccatcgatc 5916
Db 5674 cagttgtacaaagccacccaaagcattgcacacgcaacacaaatgagcttgaccatcgatc 5733
QY 5917 caccacaaacaaataaagcgaatgagggatttccatcagcagcagcagcagcagcagcagc 5976
Db 5734 caccacaaacaaataaagcgaatgagggatttccatcagcagcagcagcagcagcagcagc 5793
QY 5977 ccacaaagcctacattcctgcagatccattggttaccgggggtgtgtgtgtgtgtgtgtgtgt 6036
Db 5794 ccacaaagcctacattcctgcagatccattggttaccgggggtgtgtgtgtgtgtgtgtgtgt 5853
QY 6037 caagtgccgtgtgagcgtgggactgtcgaagcgtgtcgaatgtcgaatgtcgaatgtcgaat 6096
Db 5854 caagtgccgtgtgagcgtgggactgtcgaagcgtgtcgaatgtcgaatgtcgaatgtcgaat 5913
QY 6097 atcaatgttcagccgataccacaaagccatgtagggcgagtggtgtgtgtgtgtgtgtgtgt 6156
Db 5914 atcaatgttcagccgataccacaaagccatgtagggcgagtggtgtgtgtgtgtgtgtgtgt 5973
QY 6157 ttt 6159
Db 5974 ttt 5976
```

RESULT 8  
AAF59106  
ID AAF59106 standard; DNA; 6259 BP.  
XX  
AC AAF59106;

```

XX 24-APR-2001 (first entry)
DT
XX M. catarrhalis M56 200kDa gene in pKS348 SEQ ID NO:12.
DE
XX Moraxella catarrhalis strain Q8; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection; ds.
XX
XX Moraxella catarrhalis.
OS
XX WO200107619-A1.
PN
XX 01-FEB-2001.
PD
XX 26-JUL-2000; 2000WO-CA00870.
PF
XX 27-JUL-1999; 99US-0361619.
PR
XX (CONN-) CONNAUGHT LAB LTD.
PA
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;
PI
XX WPI: 2001-159722/16.
XX P-PSDB: AAB69127.
DR
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis
XX
XX Claim 1; Fig 8A-V; 247pp; English.
XX
XX The present invention describes an isolated and purified nucleic acid (I)
XX that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
XX The 200 kDa outer membrane protein (II) has antibacterial activity and
XX can be used in vaccines. (II), and its truncated versions, are used as
XX immunogenic compositions and vaccines to protect against M. catarrhalis
XX infections, particularly otitis media in humans. (II) is also used as
XX antigen in immunoassays for detecting specific antibodies (Ab), and to
XX generate Ab. (I) are used for recombinant production of (II) and its
XX fragments are used as probes for identifying/cloning 200 kDa protein
XX from other strains, and for diagnostic detection of M. catarrhalis.
XX (I) makes possible production of large amount of recombinant immunogens.
XX Expression of truncated versions of (II) reduces toxicity of the protein
XX towards the Escherichia coli host. The present sequence represents the
XX M. catarrhalis M56 200kDa gene in pKS348, which is given in the
XX exemplification of the present invention.
XX
XX Sequence 6259 BP; 2067 A; 1414 C; 1393 G; 1385 T; 0 other;
XX SQ
Query Match 64.1%; Score 3946.8; DB 22; Length 6259;
Best Local Similarity 80.9%; Pred. No. 0;
Matches 4904; Conservative 0; Mismatches 1002; Indels 156; Gaps 20;
QY 167 tgatcggtgcgagcgtcaatggcagtgcttatgtctcaacaaattactacaaag---atcg 223
Db 2 tgatcggtgcagcgtcagtggtgcttatgtctcaacaaattactacaaacataatcg 61
QY 224 aaattggtcaacaaacaaagataaacacacgcgtgaagggcgatgccctagcgacaggtcg 283
Db 62 caattggtgaacaaacacacagcgaagacgtc-----aggcactgccagggcgagcgtg 115
QY 284 aagcatccatctgttttggtagtcttcttaaggcacaaggtctcgaagcttctgctatcg 343
Db 116 atcgagccatctgtattgtgaaatgctaaacgcacggggcggtcgaagccatcgccatcg 175
QY 344 gtagtgcataaacacagatcctaataatgtagtaatgtagttagtgcacacgcaaaag 403
Db 176 gtagtgcataaacacactgtcctaataatgtagtaatgtagttagtgcacacgcaaaag 235
QY 404 gtaacgagtcctcatcgccatcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 463
Db 236 gtcaagagtcctcatcgccatcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 295
```

Qy	1499	aaagcaaaccttactcttgacaacgaaaagcgtataaagttggcaacagcacccctaaacagtg	1558
Db	1499		1558
Db	1349	aaacagcaccatacttggtataaaacaaacttaaagttgagttgtgcaattaccatag	1408
Qy	1559	gtagctgactgttaataacacaccactggtaataaacaacaaatccaagtcgggtgctaattggca	1618
Db	1409	acaaatggcatgtgacgagtgataaaaaagatcagtaactctgccaadgggtagcagtgcta	1468
Qy	1619	ttaaattgccaacagtcgcgttaataatgtttgcaaatcacctcagcaacagtcgggcactgctc	1678
Db	1469	acgatcggtttaccatcgaacagcgtcaaaagccgcaagccctactttaaagcagcgctg	1528
Qy	1679	gtattaccgaagaaaaattggtttttgctgtgtactaactgatggaagttgatgaacaagcac	1738
Db	1529	gcatacagttcacacactactgaaatactcagttcagttgatgctaaagttggcaattaccgccc	1588
Qy	1739	catatttgataaagaacgactaaagtgggtcgtgttgaaattacacacagatagtgtta	1798
Db	1589	caacttacaaca-----tggcgtgaaacaccaccgagcttaacagtgatggcacta	1639
Qy	1799	ttaatgctgttaatacacaagattaccggacttactaaattgtatgataagcaaataccgatcgg	1858
Db	1640	gtataaaatttagtttaagggtagttgtacgaacaatactgatttagttaccgcgcaacatt	1699
Qy	1859	ttaccataaacagcgtcaaaagacgccaagcctactttaaacgcgaggcgatggcatcagta	1918
Db	1700	tggcaagctatcataatgaagtcacatcgaaacgctgacagtgctctacaagctttacg	1759
Qy	1919	ttaatagtaataacggggatctagttgatagtagtggcaattattaccaccccaacttata	1978
Db	1760	ttaaagaagaagcatgatgaag-----ccaacg	1789
Qy	1979	acattagcttgaaacccacttaagcttaacagtaattggcaccagtggttaataataattta	2038
Db	1790	ctatcacctggctaaagatacagacaaaaatggcggcgagtcagcatcttaaaactca	1849
Qy	2039	gtgttagtaatgcctatgataacaatagcttagttaccgccaagatttggcagactatc	2098
Db	1850	aaggtaaacacggtctaaacggttgctaccaaaaaagatggtaaggttacctttgggtta	1909
Qy	2099	taaataaagtcataaagcgtgacagtgctctaccacgctttaaagtcacaaacggtg	2158
Db	1910	gcaagatagcgtctcgacatttggcaaaagcaccttaacaaacgattgcttgactgta	1969
Qy	2159	ataatgcaacaacgcctcacccgtgggtcaaaagatacaaaacggaagccttcacacct	2218
Db	1970	aagataccaacgaacaaatccaagtcgggtgctaaatggcatataatttactaatgtgaatg	2029
Qy	2219	taaaactcaaaagtgaaaacggttggttaattattaccaccaatagagccacaggtacagta	2278
Db	2030	gtagtaatccaggtactggcatttgcaaataccgctcgcatata---ccaagataaaattg	2086
Qy	2279	cctttggcatgtgcaaaagtaatggtctacacgccttaagctgacctgggttagcgata	2338
Db	2087	gctttgctgttctgatgtgcaattgatac-----aaacaacaccttacttgataca	2139
Qy	2339	caaatggttaatcgatggtttatttagcaagtcacctagcgtgacggttaacagcaccaaaa	2398
Db	2140	gacaagcta--caagttggcaatgttaagattaccaaacactggcatcaacgaggttgta	2197
Qy	2399	acatcataaaggattgtcccaacactgccttagctatgctgcagtcgcaagtgccgcgaaca	2458
Db	2198	aagccatacagggctgtcccaacactgccttagcatgctgacgtcaaaagtagcgcaaca	2257
Qy	2459	tagcactgggcaatacaaatcgaagaaaaagacaaatccaacgctgcgaagcttagatgatg	2518
Db	2258	tagaactgggcaatacaatccaagacaaagacaaatccaacgctgcgaagcttaagtata	2317
Qy	2519	tgctaaatcagagctttaacctaaaaataatgtggcaagacaaagaaacttgtctccactt	2578
Db	2318	tattcaatacaggtcttaacctaaaaataataaacaacccactgactgtgtgtcccaact	2377
Qy	2579	atgacactgttgactttatcgatggcaatgccacacacgcgcacagtaacttatgataaag	2638

[illegible]









Qy 1996 actaagcttaacagtaaatgcccacagtggttaataataaatttagtgttagtaattgctcat 2055  
Db 1953 ----- 1952  
Qy 2056 gataacaatagcttagttaccgccaaagatttgcagactactataaataaagtcaatgaa 2115  
Db 1953 ----- 1952  
Qy 2116 acgggtgacagtgctctaccgaagctttaaagtcacaaacggtgataatagcaacaacgcc 2175  
Db 1953 ----- 1952  
Qy 2176 atcaccgtgggtaaaagatacaaaaacggaagaccttcaaacctttaaacactcaaaaggtgaa 2235  
Db 1953 ----- 1952  
Qy 2236 aacggtgttaataatctagaccataatagaccacaggtacagttaccttggcattgaccaa 2295  
Db 2002 aacggatcagtaicagcaaatgacatagccaaaggtaaagtcacaaagttggtatgaccca 2061  
Qy 2296 agtaatggtctacaccgcctaaagctgaccgtgggtgagcgatatacaaaatggtaa ---tcga 2352  
Db 2062 atcaatggtctcaccgcctaaagctgaccgtgggtgagcgataaagattgtaaaactcaa 2121  
Qy 2353 ttggttattgagcaagtccttagcgtgacggttaacagcgaccacaaacacataattaaagga 2412  
Db 2122 ttggttattgagcaagttggttag -----cggtaacgacacacaaacacatacttagagga 2175  
Qy 2413 ttgtcccaacactgcctagctattgcccagtcacagttgc ---cgcaacatagcactgggc 2469  
Db 2176 ttgtcccaacactgcctagctattcacaatgaggtgggtgacgcacacaaagacgggc 2235  
Qy 2470 aatacaactcgaa ---gaaaaagacaaatccaaacgctgcacgattgtagtgtctaaat 2526  
Db 2236 aatacaactcaccgacgacgaagacaaatccaaacgctgcacgattcgggtgatatataat 2295  
Qy 2527 gcaggcttacccttaaaaaataatgggaagacaaagactttgtctcaacttagacact 2586  
Db 2296 acaggcttacccttaaaaaataatgacaaactccgttgggttgcctcaactataaact 2355  
Qy 2587 gttgactttatcgatgcaatgccacacgcacagtaacttatgatgaagccaatacaa 2646  
Db 2356 gttgactttatcgatgcaatggcaccacgcctaaagtaacttcagatgaaaccaaata 2415  
Qy 2647 accgtaagtgccgtatgattgaattggtgagagaaacacattgaactgacagggcgat 2706  
Db 2416 accgtaagtaacttatgattgaattggtgagagaaacacattgaactcacagggcgat 2475  
Qy 2707 aatggcaag ---aaacacttggcgtaacaaacacatcaaaactgacggaaacaaagtactaat 2763  
Db 2476 aatggcaagcaaaataatggcgtaacaaacacacactgacccaacaaatgctaat 2535  
Qy 2764 ggtatgcaactacatttagtaccgacgtagacacatgccttgtttaaagccagtgatgc 2823  
Db 2536 ggtaaagcaacaaactttagtaccacgataacagatgcccctgtttaaagccaaagacatc 2595  
Qy 2824 gccggcaacttaaacacacttagcggagaaattcacacacacaaagggacagcaaacacc 2883  
Db 2596 gccgaaatcLaaacacacttagcgaagaaattcacacacacaaagggacagcagacacc 2655  
Qy 2884 gccctacaaactttaccgtttaaaaggttagatgaaatgataagggctgtagcaccaaac 2943  
Db 2656 gccctacaaactttaaagtcacaaagagcgtgcaactgata -----cgaa 2703  
Qy 2944 gccatcccggtggtaagatggcaca -----agtggttaagtcacacacttaaaactc 2997  
Db 2704 accatcccggtgggttaagatgggtacacaaacacggcgaagaccgtcaacactctaaactc 2763  
Qy 2998 aaaggtaaaaacgcttctgtattataaacaccgacaaagatggtacggtttacctttggcatt 3057  
Db 2764 aaaggtgaaaacgcttcaacggttgctaccacataaagatggtacggtttacctttggcatt 2823

Qy 3058 aacacccaaaacggtctttaaaagccggcgacagcaccactctaaacaaacaaatggttgtct 3117  
Db 2824 aacacccaaaacggtctttaaaagccggcgacagcaccactctaaacaaagatggttgtct 2883  
Qy 3118 attaaacacccgctagtaaacgaaacaaatccaagtcggtgctgtagtgcgtgaagtttgc 3177  
Db 2884 attaaacacccgctagtaaacgaaacaaatccaagtcggtgctgtagtgcgtgaagtttgc 2943  
Qy 3178 atggttaataatggtgtttaggtgctggtgattgatggcacaacactcgcattaccagagat 3237  
Db 2944 aaggttgataaggttaattcaagcactggtgattgatggcacaagcgtatcaccaagat 3003  
Qy 3238 gaaattggcttactgggactaaatggctcacttgataaaaaacacacccactaaagcaaa 3297  
Db 3004 caaattggcttactgggctaaatggctcacttgataccacacacccactaaacaaa 3063  
Qy 3298 gac -----ggcattaaacgaggtggttaa 3321  
Db 3064 gacaagcttaagtggtgaaagtgaagtataacacacactggcattaaacgaggtggttaa 3123  
Qy 3322 aagattaccaaactcaatcaggtgagattgccaaaaaacagccactgactgctgtgacaggc 3381  
Db 3124 aagattaccaaactcaatcaggtgagattgccaaaaaacagcaatgactgctgtgacaggc 3183  
Qy 3382 ggcagaattttagtttaaaacacggaactgaaaaataaaatcagcagtagtgcgaacaa 3441  
Db 3184 ggtcgggttttagtttaaaacacggaactgaaaaataaaatcagcagtagtgcgttaaaca 3243  
Qy 3442 gcacaaactcattacacgaattctcagtagcagatgacaggtgaataaactttacgatt 3501  
Db 3244 gcacaaactcattacacgaattctcagtagcagatgacaggtgaataaactttacgatt 3303  
Qy 3502 agtaaccccttactcagttatgacacctcaaaagacactctgactgctacaccttttcaggt 3561  
Db 3304 agtaaccccttactcagttatgacacctcaaaagacactctgactgctacaccttttcaggt 3363  
Qy 3562 gaaaacggcattaccaccaaaggtaaataaaggtggtgctggtgggcatlgaccaaacc 3621  
Db 3364 gaaaacggcattaccaccaaaggtaaataaaggtggtgctggtgggcatlgaccaaacc 3423  
Qy 3622 aaagcttaacacgcctaaagctacgctgggtgaataataatggcagaagcatttgcatt 3681  
Db 3424 aaagcttaacacgcctaaagctacgctgggtgaataataatggcagaagcatttgcatt 3483  
Qy 3682 aacagcgaataatggtcaaaataacacacagcagtaacgaacactctgacttaatttacc 3741  
Db 3484 gacagtaaaagtgtcaaaataacacacagcagtaacgaacactctgacttaatttacc 3543  
Qy 3742 aatgataaaggttagcgtacgcacacacagagagggcaataataatacaagacgaagacaaa 3801  
Db 3544 aatga -----tggtgcaggacgcagcaactaagcgaagggcttgccaatgacacgcacaaa 3597  
Qy 3802 accgctgcgcagcagcattgttgatgctaaagcagcagcttttaacttgcaagggcaatggt 3861  
Db 3598 accgctgcgcagcagcattgttgatgctaaacgcagccttttaacttgcaagggcaatggt 3657  
Qy 3862 gaagcgttgcatttgcctccacttatgacacccgtcaactttgccaatggcaataaccacc 3921  
Db 3658 gaagcgttgcatttgcctccacttatgacacccgttatgactgttgcattgcagtgccaatgccacc 3717  
Qy 3922 accgtaaggtgacctatgtagacacaaagcaaacacagtaaaagtgggttatgattgcaat 3981  
Db 3718 accgtaaggtgacctatgtagacacaaagcaaacacagtaaaagtgggttatgattgcaat 3777  
Qy 3982 gtgtagatacaacacattgaagttaaa ---gataaaaaacttggcgtaaaacacccacaa 4038  
Db 3778 gtggaataaaaaacacattgaagtgacagtgataaaaaaacttggcgtaaaacacccacaa 3837  
Qy 4039 ttgacagcagtaggcacaggtgctaaataatttgccttaagcaatacaagctactgctgcat 4098  
Db 3838 ctgacacaaaaacaaagtgctaaatggtaatgca ---accaaatttggcgcgatggcgat 3894  
Qy 4099 gcgcttgtaagggccagtgatgcattgctcattctaaacacacacttcttggcgacatccaa 4158

[illegible]

```
Qy 5332 cagcccaaggcagatggtgaagccgcgcgttgccatagggcagacaaacccaagcaggcaac 5391
Dy 6115 cagcccaaggcagatggtgaagccgcgcgttgccatagggcagacaaacccaagcaggcaac 6174
Qy 5392 caatccatcgccatcggtgataacgcgaacgcagccagcgatcaatccatcgccatcggt 5451
Dy 6175 caatccatcgccatcggtgataacgcgaacgcagccagcgatcaatccatcgccatcggt 6234
Qy 5452 acaggcaatggtgtagcaggtgaagcactctggtgccatcgccgagcccaagcactgttaag 5511
Dy 6235 acaggcaatggtgtagcaggtgaagcactctggtgccatcgccgagcccaagcactgttaag 6294
Qy 5512 gctgataaacagttacagttggtgataaacacacagtttaccgattgccaactcaaacagat 5571
Dy 6295 gctgataaacagttacagttggtgataaacacacagtttaccgattgccaactcaaacagat 6354
Qy 5572 gctcttggtggtggaataacatcacctgacccggtgacgaagtaactcggttgccttaggttca 5631
Dy 6355 gctcttggtggtggaataacatcacctgacccggtgacgaagtaactcggttgccttaggttca 6414
Qy 5632 aactctgccatcagtcagggcacacagcagggcacacaaagccaaatactgacggcaca 5691
Dy 6415 aactctgccatcagtcagggcacacagcagggcacacaaagccaaatactgacggcaca 6474
Qy 5692 gcaggtacaacacacacagcaggtgcccacaggtgacggttaaaagcgttctgctggacaacg 5751
Dy 6475 gcaggtacaacacacacagcaggtgcccacaggtgacggttaaaagcgttctgctggacaacg 6534
Qy 5752 gcggttggtgcgtctccgtggtggtgctcaggtgctgaacgcgcctatccaaaatgtggca 5811
Dy 6535 gcggttggtgcgtctccgtggtggtgctcaggtgctgaacgcgcctatccaaaatgtggca 6594
Qy 5812 gcaggtgaggtcagtcagccacagcagcagcagtcggtcaatggtgtagcagttgacaaagcc 5871
Dy 6595 gcaggtgaggtcagtcagccacagcagcagcagtcggtcaatggtgtagcagttgacaaagcc 6654
Qy 5872 acccaagcattgcacgcgaacccaatgagcttgacatcgatccacccaaacgaacaaat 5931
Dy 6655 acccaagcattgcacgcgaacccaatgagcttgacatcgatccacccaaacgaacaaat 6714
Qy 5932 aaagccaatgcagggatttcacatgcagtcggtgagtcggtcgcacacagcgcctacatt 5991
Dy 6715 aaagccaatgcagggatttcacatgcagtcggtgagtcggtcgcacacagcgcctacatt 6774
Qy 5992 cctggcagatccatggttacccgggggtattgccaccacacacacagtcgaaggtgcggtggca 6051
Dy 6775 cctggcagatccatggttacccgggggtattgccaccacacacacagtcgaaggtgcggtggca 6834
Qy 6052 gtgggactgtcgaagctgcgggataatggtcaatggtgatttataaatacaatggttcagcc 6111
Dy 6835 gtgggactgtcgaagctgcgggataatggtcaatggtgatttataaatacaatggttcagcc 6894
Qy 6112 gatacccaagccatgttagggcgccagttggtgtagcaggttttcacttt 6159
Dy 6895 gatacccaagccatgttagggcgccagttggtgtagcaggttttcacttt 6942
```

## RESULT 10

AAF59129

ID AAF59129 standard; DNA; 720 BP.

XX

AC AAF59129;

XX

DT 24-APR-2001 (first entry)

XX

DE M. catarrhalis strain 4223 200kDa partial nucleotide sequence.

XX

KW Moraxella catarrhalis strain 4223; major outer membrane protein;  
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection; ds.

XX

OS Moraxella catarrhalis.

XX WO200107619-A1.

PN

XX 01-FEB-2001.

PD

XX 26-JUL-2000; 2000WO-CA00870.

PF

XX 27-JUL-1999; 99US-0361619.

XX

XX (CONN-) CONNAUGHT LAB LTD.

PA

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

PI

XX WPI; 2001-159722/16.

XX

DR P-PSDB; AAB69147.

DR

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

PT useful in protective vaccines and for diagnosis

PT

XX Example 14; Fig 17; 247pp; English.

XX

The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II), and its truncated versions, are used as immunogenic compositions and vaccines to protect against M. catarrhalis infections, particularly otitis media in humans. (II) is also used as antigen in immunoassays for detecting specific antibodies (Ab), and to generate Ab. (I) are used for recombinant production of (II) and its fragments are used as probes for identifying/cloning 200 kDa protein genes from other strains, and for diagnostic detection of M. catarrhalis. (I) makes possible production of large amount of recombinant immunogens. Expression of truncated versions of (II) reduces toxicity of the protein towards the Escherichia coli host. The present sequence represents a M. catarrhalis strain 4223 200kDa partial nucleotide sequence, which is used in an example from the present invention.

XX Sequence 720 BP; 201 A; 142 C; 140 G; 237 T; 0 other;

## Query Match

Best Local Similarity 2.7%; Score 166.2; DB 22; Length 720;

Matches 171; Conservative 95.5%; Pred. No. 2.2e-33;

Mismatches 0; Indels 8; Gaps 0;

Qy 1 atgaatcacatctataaaagtcattctttaacaaagccacagggcacatttatggcgtggcg 60

Db 542 atgaatcacatctataaaagtcattctttaacaaagccacagggcacatttatggcgtggcg 601

Qy 61 gaatatgccaaatccacacagtcacgggggggggtagctgtgtacacagggcaagttggcagt 120

Db 602 gagtacgccaatccacacagtcacgggggggggtagctgtgtacacagggcaagttggcagt 661

Qy 121 gtagcactctaaagctttgcccgtattgcccgcctcgtctcctcgtatcgatcggtcgac 179

Db 662 gtagcactctgagcttgcgcgtattgcccgcctcgtctcctcgtatcgatcggtcgac 720

## RESULT 11

AAA92496

ID AAA92496 standard; DNA; 3030 BP.

XX

AC AAA92496;

XX

DT 17-JAN-2001 (first entry)

XX

DE Haemophilus influenzae adhesin (Hia) gene from NTHi strain M407.

XX

KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;  
KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;  
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;  
KW diagnosis; immunogenic; antigen; ds.

XX

OS Haemophilus influenzae.

QY	5677	aaatctgacgcgcacagcagggtacaaacacacacagcaggtgcccacagttacggtttaaagc	5736
Db	2540	aaatctgattggtacgcgcgggtaacactaataactgtgcgcacaggttacggttaaagcgc	2599
QY	5737	tttgcctggacaaacgcgcggttggtgcggtctccgtggggtcctcaggtgctgaacgcgt	5796
Db	2600	tttgcgcggtgcaacgcgcacggtgttctgtcgcgcaagtgcgaagaagaacgt	2650
QY	5797	atccaaaatgtggcagcaggttgaggtcagtgccaccacagccagtgcggtccaatggtatgc	5856
Db	2660	atccaaaacgtcgcgcagcgcgaattccgcgtacttccacagatgcgattaacggtatgc	2719
QY	5857	cagttgtacaaagccaccacaaagcatgtccaaagcaacaaatgaggttgaccatcgtatc	5916
Db	2720	cagttgtatcgtgtggcaaaaggggtaacaaaac-----ctgtgtgacaaagt	2767
QY	5917	caccaaaacgaaataaaagccaatgcagggaattctatcagcgaatggcgaatggcggtccatg	5976
Db	2768	aataaagtgggcaaacgtgcagatgcaggtgcagcaagtgcattagcgggttcacagtta	2827
QY	5977	ccaaagcctcatcttcctgcagatccatggttacgcggggtattgccaccaccaacggt	6036
Db	2828	ccaaagcctcta tgcaggtaataacatggttctctatttcggggaagttagttatcaaggt	2887
QY	6037	caaggtgcggtggcagtgggagctgcgaagctgtcggataatggtccaatgggtattaaa	6096
Db	2888	caagatggttttagcta tccgggtatccaagaatttcgga taatggcagaagttattatgcg	2947

5736

Db 2555 aaatctgtagtcgagcggtgaaacactacaactgctgtgcacaaacgggacggttaaacggc 2614  
 Qy 5737 ttctgtgacaaaacggcggttggtgctgtccctggtggtgctcaggtgctggaacgcgt 5796  
 Db 2615 ttgctggtgcaacggcgacggtgctgtctcgtcgcgcaagcgcggaagaaacgt 2674  
 Qy 5797 atccaaaatgtggcagcaggtgaagtcagtgccaccacgacccgagtcggtcaatggttagc 5856  
 Db 2675 atccaaaacgttggcgagcggaatttccgctacttccaccgagtcgattaaacggcagc 2734  
 Qy 5857 cagttgtacaagccaccaaagcattgccaaacgcaacaaatgagctgaccatcgatc 5916  
 Db 2735 cagttgtatgcgtggcaaaagggttaacaaac-----cttgcgtggacaagt 2782  
 Qy 5917 caccaaaacgaaaataaaagccaatgcaggggatttccatcagcgatggcggtcccatg 5976  
 Db 2783 aataaagtggcaaaacgtgcagatgcaggtacagcaagtcattagcgttccacagtta 2842  
 Qy 5977 ccacaagcctacattcctggcagatccatggtttaccgggggtattgccaccacaaacggt 6036  
 Db 2843 ccacaagcctctatgtcaggttaaatcaatggtttctatttgcgggaagttagttatcaagt 2902  
 Qy 6037 caaggtgcgtggcagtgagctgcgaagctgcgagataatggtcgaatggttatttaa 6096  
 Db 2903 caaagtggttagctatcggtgtacaaagatttccgataatggaagtgcaaatgattattcgc 2962  
 Qy 6097 atcaatggttcagcgataccacagggcagtgtagggcgcgagttggtgcaggttttccac 6156  
 Db 2963 ttgtcaggcacaaacaaatagccaaggttaaacacagcggttcagcaggtgttggttaccag 3022  
 Qy 6157 t 6157  
 Db 3023 t 3023

## RESULT 13

ID AAA92495 standard; DNA; 3354 BP.  
 XX  
 AC AAA92495;

17-JAN-2001 (first entry)

Haemophilus influenzae adhesin (Hia) gene from NTHi strain K9.

XX Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;  
 KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;  
 KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;  
 KW diagnosis; immunogenic; antigen; ds.

Haemophilus influenzae.

WO200055191-A2.

21-SEP-2000.

16-MAR-2000; 2000WO-CA00289.

16-MAR-1999; 99US-0268347.

(CONN-) CONNAUGHT LAB LTD.

Loosmore SM, Yang Y, Klein MH;

WPI; 2000-618897/59.

P-PSDB; AAB23858.

PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for  
 PT use as antigens and vaccines and for treating Hemophilus influenzae  
 PT infection

PS Claim 1; Fig 22; 275pp; English.

XX

CC The present sequence represents a Haemophilus influenzae adhesin (Hia)  
 CC gene from the non-typeable Haemophilus influenzae (NTHi) strain K9.  
 CC Hia genes and proteins have antiinflammatory, auditory and antibacterial  
 CC activities, and can be used in the production of a vaccine. An  
 CC immunogenic composition comprising an Hia gene, a polypeptide encoded  
 CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
 CC protection against disease caused by Haemophilus strains in a  
 CC susceptible host, preferably a human. An Hia protein is useful as an  
 CC antigen, in immunogenic preparations including vaccines, as a carrier  
 CC for other immunogens, and in the generation of diagnostic reagents. Hia  
 CC is useful for treating diseases caused by the infection of Haemophilus  
 CC influenzae such as meningitis, epiglottitis, septicemia and otitis  
 CC media. Recombinant production of Hia favours high recovery of the  
 CC protein compared to the low recovery of native protein from Haemophilus  
 CC influenzae species. A truncated protein has a significantly higher  
 CC amount of recovery than a full-length protein.

XX Sequence 3354 BP; 1082 A; 639 C; 866 G; 767 T; 0 other;

Query Match 2.2%; Score 138.2; DB 21; Length 3354;  
 Best Local Similarity 57.4%; Pred. No. 1.1e-25;  
 Matches 276; Conservative 0; Mismatches 193; Indels 12; Gaps 1;

Qy 5677 aaatctgacggcagcaggtgtacaaccacacgaggtgcccacaggttacgtttaaagc 5736

Db 2864 aaatctgattgacggcggttaacactacaactgtgcacacggtgtacgttaacg 2923

Qy 5737 ttgtgtgacaaaacgggtgtgctgtcgttcgtggtggtgctcaggtgtcgtgaaacg 5796

Db 2924 ttgtcgtgtcaacggcgacggtgtcgtgttctgtcgtcgcaagcggaagaagc 2983

Qy 5797 atccaaaatgtggcagcaggtgaggtcagtcgacacgacgacgagtcggtcaatgtagc 5856

Db 2984 atccaaaacgtcgcgagcgaggaatttccgcaattccacgagtcgattaacggcagc 3043

Qy 5857 cagttgtacaagccacccaaagcattgccaacgcaaccaatgagcttgaccatcgatc 5916

Db 3044 cagttgtatgctg-----tggcaaaagggttaacaaatcttctgtggaacagtg 3091

Qy 5917 caccaaaacgaaaataaaagccaatgcagggatttcatcagcagatggcggtccatg 5976

Db 3092 aataaagtggcacaacgtgcagatgcaggtacagcaagtcattagcagcttcacagtta 3151

Qy 5977 ccacaagcctacattcctgcagatccatggttacgggggtattgccaccacacacgt 6036

Db 3152 ccacaagcctctatccaggttaatacaatggtttctattgcgggaagtagttatcaaggt 3211

Qy 6037 caaggtgcgttggcagtggtgagctgtcgaagctgtcggataatggttcaatgggtatttaa 6096

Db 3212 caaaatggttagctatcggtgtatcaacgaatttccgataatggcgaagtgattattcgc 3271

Qy 6097 atcaatggttcagccgatacccaagccatgtaggggcgcgaggtgtggtcaggttttccac 6156

Db 3272 ttgtcagggcacaacaaatagccaaggttaaacacaggtgtgacaggtgtgtgtaccag 3331

Qy 6157 t 6157

Db 3332 t 3332

## RESULT 14

AAA92498

ID AAA92498 standard; DNA; 3342 BP.

XX

AC AAA92498;

XX

DT 17-JAN-2001 (first entry)

XX Haemophilus influenzae adhesin (Hia) gene from NTHi strain K22.

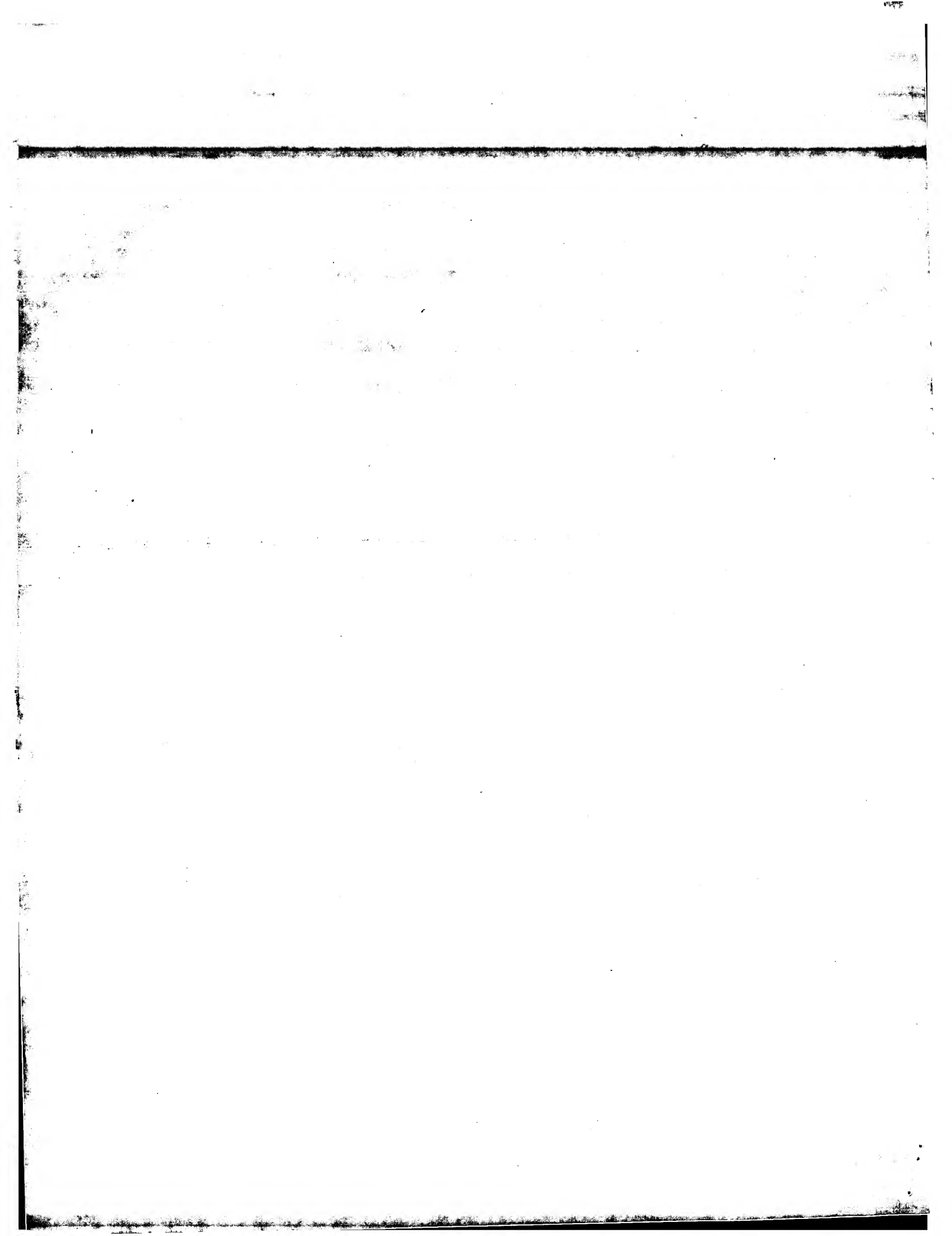
DE

XX Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;

[illegible]



Search completed: September 12, 2001, 13:39:40  
Job time: 3602 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 12:54:07 ; Search time 454.28 Seconds  
(without alignments)  
2566.627 Million cell updates/sec

Title: US-09-361-619-8  
Perfect score: 6159  
Sequence: 1 atgaatcacatctataaagt.....ttggtgcagggtttcacttt 6159

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4090.6	66.4	6973	1	US-08-478-370-1
2	4001.4	65.0	9542	4	US-08-968-685A-9
3	84.4	1.4	3294	1	US-08-409-995-1
4	84.4	1.4	3294	3	US-08-685-467-1
5	84.4	1.4	3294	4	US-08-913-942-1
6	83.6	1.4	2037	4	US-08-913-942-14
7	77.2	1.3	7291	4	US-08-913-942-3
8	66	1.1	5738	1	US-08-409-995-3
9	66	1.1	5738	3	US-08-685-467-3
10	60	1.0	60	1	US-08-478-370-4
11	58.4	0.9	1797	4	US-09-377-155-12
12	56.8	0.9	1797	4	US-09-377-155-4
13	55.2	0.9	1770	4	US-09-377-155-18
14	55.2	0.9	1776	4	US-09-377-155-10
15	55.2	0.9	1776	4	US-09-377-155-20
16	55.2	0.9	1779	4	US-09-377-155-3
17	55.2	0.9	2308	4	US-09-377-155-1
18	53.8	0.9	1718	1	US-08-232-463-14
19	53.6	0.9	1785	4	US-09-377-155-6
20	53.6	0.9	1785	4	US-09-377-155-8
21	53.6	0.9	1800	4	US-09-377-155-14
22	52	0.8	1779	4	US-09-377-155-16
23	48.4	0.8	414	2	US-08-630-822A-63
24	48.4	0.8	414	2	US-09-005-069-63
25	48.2	0.8	2277	1	US-08-676-967-2
26	48.2	0.8	2277	1	US-08-676-974-2
27	48.2	0.8	2277	2	US-09-098-487-2

28	45.8	0.7	4765	1	US-08-750-532-8	Sequence 8, Appli
29	43.6	0.7	3489	2	US-08-728-323A-1	Sequence 1, Appli
30	43.6	0.7	32207	2	US-08-770-379-20	Sequence 20, Appl
31	43.6	0.7	32207	4	US-08-757-669A-20	Sequence 23, Appl
32	42.2	0.7	15222	2	US-08-801-898A-23	Sequence 12, Appl
33	42.2	0.7	15222	4	US-08-962-690-12	Sequence 1, Appli
34	42.2	0.7	15223	2	US-08-892-403A-1	Sequence 65, Appl
35	41.4	0.7	273	2	US-08-630-822A-65	Sequence 65, Appl
36	41.4	0.7	273	2	US-09-005-069-65	Sequence 71, Appl
37	40.6	0.7	293	2	US-08-630-822A-71	Sequence 71, Appl
38	40.6	0.7	293	2	US-09-005-069-71	Sequence 18, Appl
39	40	0.6	615	4	US-08-998-416-186	Sequence 186, App
40	39.2	0.6	819	1	US-08-320-161-3	Sequence 3, Appli
41	39.2	0.6	819	3	US-08-642-807A-18	Sequence 3, Appli
42	39.2	0.6	819	4	US-08-455-829-3	Sequence 69, Appl
43	39.2	0.6	1991	4	US-08-235-836C-69	Sequence 77, Appl
44	39.2	0.6	1991	4	US-08-235-836C-77	Sequence 41, Appl
45	39	0.6	1462	3	US-08-961-083-41	

ALIGNMENTS

RESULT 1  
US-08-478-370-1  
; Sequence 1, Application US/08478370  
; Patent No. 5808024  
; GENERAL INFORMATION:  
; APPLICANT: SASAKI, Ken  
; APPLICANT: HARKNESS, Robin E.  
; APPLICANT: LOOSMORE, Sheena M.  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER  
; MEMBRANE PROTEIN OF MORAXELLA  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,370  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-502  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6973 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-478-370-1

Query Match 66.4%; Score 4090.6; DB 1; Length 6973;  
Best Local Similarity 81.3%; Pred. No. 0;  
Matches 5063; Conservative 0; Mismatches 1009; Indels 157; Gaps 21;  
Oy 1 atgaatcacatctataaagtctattttaaagccacaggcacatttatgcccgtggcg 60  
|||||

D	b	542	ATGAATCACATCTATAAAGTCATCTTTAACAAAGCCACAGGCACAATTATGTGGCAGTGGCA	601
Q	y	61	gaatattgccaataccccacagtagc-gggggggggtagctgtctacaaggccaagtgtgcag	119
D	b	602	GAGTAGCCCAANTCCACACGACGCGGGGGGTAGCTGTGTACAGGGCAAGTTGGCAG	661
Q	y	120	tgtacgcactctaagcttttgcccgatttcgcgcgcctcgtcctcgtgatcggtgcgac	179
D	b	662	TGTAATGCACACTGAGCTTTGCCCGTATTGCCGGCGCTGCTGCTCTGTCGTATCGGTGCAAC	721
Q	y	180	gctcaatggcagtgcttatgtctcaacaaattactacca---agatcgaaaattggttcaaac	236
D	b	722	GCTCAGTGGCAGTGCTTATGCTCAAAAAAAGATACCAACAATATCGCAATTGSTGAACA	781
Q	y	237	aaacaagaataacaacacgcgtgaaaggcagtcacctagcacagcagtgtaagcatccattgc	296
D	b	782	MAACGAGCAAGACGCTC-----AGCAGCTGCCAAGGGCGGTGATCGAGGCCATTGC	835
Q	y	297	tttgttagtctttctaaaggcacaggctctcaagctattgttatcgttgtagtgcataacc	356
D	b	836	TATTGGTGAAAATGCTTAACGCACAGGGCGGTCAAGGCATCGCCATCGGTAGTAGTAATAA	895
Q	y	357	agatcctcataatggtadgaatgtaatgtaggttcccacgcgaaggtaaacsagttccat	416
D	b	896	AACGTGTAATGGAACACAGTTTTGGATAGATAGTAGGTACCAGTGTACGGGTCAAGAGTCCAT	955
Q	y	417	cgccatcagtggtgatgtattggctgaggggtgatgcctcgtatgcatacoggtagtgatga	476
D	b	956	CGCATCGGTGGTATGATNAAGGCTAGTGGTATGCTCGATTGCCATCGGTAGTAGTGA	1015
Q	y	477	cttataattgctctaagaatccttgatctgaagaa-----tgeaatcttcaaa	521
D	b	1016	CTTACATTTTGCTGTATCAGCATGTAATCTTAACATCCGNAAGGTACTCTGTATTAACGA	1075
Q	y	522	actattatcgtgcataaataattaaaaaaatacaaacctcaacccagatggttaaaatcaa	581
D	b	1076	TCATTATTAAACGGCCATCGACTATTAAAGAANAATACGAAGCTCAAAGGATAATGATGTAA	1135
Q	y	582	atatcgacgcacaagcgcacaaggcgcagccagtaactcagtgagagccatgctcatatgc	641
D	b	1136	ATATPAGACGCACAACCGCAAGCGGACACGCCAGTACTGCACTGGGAGCCATGTCTATATGC	1195
Q	y	642	acaggttcatttttcaaacgcctttggtatcatcaccaacagctgaagctgctattccct	701
D	b	1196	ACAGGGTCATTTTCCNAACGCTTTGGTACAGGGCAACAGACTAAAGTGCTATTCCTT	1255
Q	y	702	ggcagtaggttttgcgcgcacacacaaacaacttcaatcgctgtgtgttccaatgc	761
D	b	1256	GGCAGTGGTCTTGCCCCCACAGCCGAGGCCCAANTCTACATCGCTATTGGTTCTGATGC	1315
Q	y	762	aaaagctaaccgcgtttgcagcgacagccattggtgaaatactgtagttaatttgggtcg	821
D	b	1316	AACATCTAGTCGTTGGNGCGATAGCCCTTGGTGCAGGTACTCTGTCFACGTACAGGG	1375
Q	y	822	agcgtgtcccttagtttgggttcagatcccttgataggataataatacacagatgccag	881
D	b	1376	CAGTATTGCCCTAGGTCAAGGTTCTGTGTGCATCAGAGGTGATPAATAATC---TAGACC	1432
Q	y	882	tgcctatgtaccactagtaaaaaagtttagcagaccagataaaagccacccgcagggtag	941
D	b	1433	GGCCTATACACCAATACCCAGGCATAGACCCCAGTTTCAAGCCAC---CAATAATATAC	1489
Q	y	942	ttctacggataattttocattggttaatagtaataataataatagcagatcatcaggcgtaa	1001
D	b	1490	GAAGGGGGTCCCATTTCCATTGTTGATGA-----CTCTATCAACAGSTAA	1534
Q	y	1002	aatcatcaaatgltcggttgcgggtttctcgggatacccgatgcggtcaatdgtggcacagcttaa	1061
D	b	1535	AATCATCATMTGCGTGCAGGTTTAATAAAAAACCGATCGCGTCAATGTGGCAGCAGTAGA	1594
Q	y	1062	attgggtggaggaaactggctca---alcgtaaaattacttttaaggggtgatggtgacaataa	1118
D	b	1595	AGCGGTGGTGAAGTGGGGCTTAAGAGCGGTAGAAATTACTTTTCAGGGTGNAG-----ATAA	1648

QY	1119	tagcaatagcgtgaaagagcttggcgaatactctttaactattaaagtgatgcacgac	1178
DB	1649	CAGTACTGACGTAAATAATGATGGTTGGATATACTTTAACTATTAAGGTGGTGCAGAGAC	1708
QY	1179	caacgattacccgaagctaaactcgtgtggtgtaacagatggcaa-----tggctgtaa	1232
DB	1709	CAACGCATTACCCGATATATATATCGGTGGTAAAGAGGCTGATTAATAGTGGTTCGAA	1768
QY	1233	gcttaaacctgtcaagagcgtgactggatggaccagtgctcc-----gc	1277
DB	1769	AGTTAAACTTCTTAAACTTTAAACAATCTTACTGAGGTGAATACAACTACATTAATATGC	1828
QY	1278	taccacaataatcacccgttagtaatacaacaacaacacgcgcgagctacaagcgttg	1337
DB	1829	CACAACCCACAGTTAAGGTAGGTAGTAGTAGTACTACAGCTGAATATTATTGAGTGATAG	1888
QY	1338	tttgacctttagccca-----ataacagggtacaaaaacagataaaacgctctacagcat	1391
DB	1889	TTTTAACTTTTACCAGCCCAATACAGGCAGGTCAAGGACACAAAGAAACCGCTATGCGGT	1948
QY	1392	tgatggattgaagtttactaatgatgaatagtagcaactaaagggtactactcgtat	1451
DB	1949	TAAATGGGGTGAAGTTTACTTAATATTCAGAAACACACAGCAGCAATCGGCACACTACTCGTAT	2008
QY	1452	tacaaaaagaaattgggttttgcgtggtactaatgtagtgaggtgatgaagacaaacctta	1511
DB	2009	TACCAGAGATAAAATTTGGCTTTGCTCGAGATGGTGATGTTGATGAANAACNAGCCACATA	2068
QY	1512	tcttgacaacgaagactaaagtttggcaacgacccctaaacagttgtagctgactgt	1571
DB	2069	TTTTGGATTAATAAACACTTTAAAGTGGGTAGTTGTCGAAGAGTACCACTGCTAACGATCGGTTAC	2128
QY	1572	taataacaccactggtaataaacaacaaatccaaatcggtgctgaatggcattaaattggcac	1631
DB	2129	TGCAGGTAAATAAAGATCATGTAATCTGCCAAAGAGTACCACTTACCATAGACAATGGCATTC	2188
QY	1632	agtcgctaataatgttgcaataacctcgaacaacgctggcagctgctgtattatcacgaaga	1691
DB	2189	CATCGACAGCTCANAGCCGCAAGGCTACTTTAAACGCGAGCGCTGGCATTGAGTGTGCAC	2248
QY	1692	gaaattgggttctggtactaatgatggattgatgaacgaagcaccattatttggataa	1751
DB	2249	ACCTACTGAAATATACGTTTATGCTAGAGTGGCAATGTTTACCGCCCAACTTTACAACA-	2307
QY	1752	agaacgacttaagtggggtcgttgtaaaatcaccaagatagtggtattaatagtcggtaa	1811
DB	2308	-----TTGGCGTGAATAACCCAGAGCTTAACAGTGATGGCAGCTAGTGATTAATTTAG	2359
QY	1812	tcacaagattaccgagcttactaatgtagcaaataccgattgcggttaccatcaaaaa	1871
DB	2360	TGTTAAGGTTAGTGTACGACANTAGCTTAGTTTACCGCGCAACTTTGGCAAGCTATCT	2419
QY	1872	gctcaaaagacgcgaagcctactttaaacgcgagcgatggcatcagattaaatagtaata	1931
DB	2420	AAATGAAGTCAATTCGAACGGGTGACAGTGCTTTCAAAAGCTTTTACCGTTTAAAGAAGA	2479
QY	1932	cggggatctagttgatagtagtggcaaatattaccaccccaacttataaacattagcgtgaa	1991
DB	2480	CGATGATGAGC-----CCAAACGCTATCACCGCTGGC	2509
QY	1992	aaccactaagcttaacagtaattggcaccagtggttaataataaatttggtgttagtaatgc	2051
DB	2510	TAAAGATACGACANAAATTTCCGGCGGACGTACGATCTTAAACTCAAAAGTAAAAACGG	2569
QY	2052	tcatgataacaaatagcttagttaccgcgcaaaagattggcagactactctaaataaagctaa	2111
DB	2570	TCTAACGGTTTGCTACCAAAAAAAGATCGTACGGTTTACCTTTTGGGCTTACGCCAAGATAGCGG	2629
QY	2112	tgaacggcttgacagtgctctaccagcttttaagtccaaaacggctgataatagcaaca	2171
DB	2630	TCTGACCATTTGGCAAAAAGACACCTTAACAACAGTGGCTTTGACTGTTAAAGATACCAACGA	2689



Db	4835	AGCCAAACCCCGAGATGGCACATTGGCTCAAAATGAATGTCAAATCAGTCAATTAACAAAGA	4894
Qy	4380	caagtaaatgatgccaataataaaagcaaggcatcaatgaagacaaacgcctttgttaaagg	4439
Db	4895	ACAAGTAATGATGCTCAATATAAAACAAGGATCAATGAAGACAACGCCCTTGTGTAAAGG	4954
Qy	4440	actgaaaaagccgctcttgataacaaacacaaacccgcgcagtaactgtgggtgattt	4499
Db	4955	ACTTTGAAAAGCCGCTTTCATACAAACCAAAACCCGCGAGTAACCTGTGGTGAATTT	5014
Qy	4500	aaatgcgctgtgccaacacgcctgacotttgcagggggatacaggcacaacgcgtataaaa	4559
Db	5015	AAATGCGGTGCCCAACACCGGTGACCTTTGACGGGATACAGGCAACAGCGCTAANA	5074
Qy	4560	actgggcgagacttgacatacaagttggcacaacagacacaataagctaaaccgataa	4619
Db	5075	ACTGGGCGAGACTTTTGACCATCAAGGTGGCAACACAGACACCAATAAGCTAACCGATAA	5134
Qy	4620	taecatcggttggttagcaggtactgatggcttcaacttcaaaacttgcaaaagacctaac	4679
Db	5135	TAACATCGTGTGTTAGCAGTACTGATGGCTCACTGTCTCAAACTTGCCAAAGACCTTAAC	5194
Qy	4680	caattctaacgcgcttaatgcaggtggccaaaactgatgaaaaggcatctcttttgt	4739
Db	5195	CAATCTTAACAGCGTTAATGCAGGTGGCACCAAAATTGATGACAAAGCGGTGCTTTTGT	5254
Qy	4740	aaagcaaacggtcaagccaaagcaaacacccctgtgtaagtgcgaatgggctggacct	4799
Db	5255	AGACTCAACGGGTCAAGCCAAAGCAAAACACCCCTGTGTCTAAGTGCAATGGCTGGACCT	5314
Qy	4800	gggtggcaaggttcatacagtaatgtggccaagcacaaaagataccgcgctgcgaatgt	4859
Db	5315	GGGTGGCAAGGTCATCAGTAATGTGGGCAAGGCAACAAAGATACCGACGCTGCCAATGT	5374
Qy	4860	acacagtttaaacgaaagtacgcacactgtttgggtct-----tggtaatgataacgc	4910
Db	5375	ACAAAGTTTAAACGAGTAGCGAACTGTTGGGTCTTGTTAATGTGTGTAAATGATAAGCG	5434
Qy	4911	tgacggcaatcagggtaaacattgcgcacatacaaaaagacccaaaactcaaggttcatatc	4970
Db	5435	TGACGGCAATCAGGTAAACATTTGCGCGATCAAAAAGACCCCAATTCAGGTTTCATCATC	5494
Qy	4971	taaccgactgtcatcaaacgcaggcaggtacttggcggtaaaggtaataacgataacga	5030
Db	5495	TAACCGCACGTCTNTCAAAGCAGGCACGTAATTTGGCGGTAAAGGTAAATACGATACCGA	5554
Qy	5031	aaaacttgcaactgggtgtacagatgggcgtgatgataagacgcgaacgctaaacggga	5090
Db	5555	AAAACTTTGCCACTGGTGGTATACAGTGGGCGTGGATTAAGACGGCAACGTAACGGCA	5614
Qy	5091	tttaagcaattgtttgggtcaaaaccccaaaagatggcagcaaaaagccctgcgtgcac	5150
Db	5615	TTTTAAGCAANTGTTTGGGTCAAAACCCCAAAAGATGGCAGCAAAAAGGCCCTGTGCCAC	5674
Qy	5151	ttataccgcgcaggttcagacacaactatgtacacaacaccccgagaagcattgacag	5210
Db	5675	TTTATACGCCCGAGGTGAGACCAACTATTTTGACCAACAAACCCCGCAAGCCATTGCACAG	5734
Qy	5211	aataaatgaacaaggtatccgctcttcacatgtcaacgatggcaatcaagagcctgtggt	5270
Db	5735	NATAATGAACNAGGTATCCGCTTCCTCATGTCAACGATGCAATCAAGAGCCTGTGTT	5794
Qy	5271	acaagcggttaacggcatctgactcaagtcctcaggccaagcacctcagtggcgatagttt	5330
Db	5795	ACAAGGCGCTAAACGGCATTTGATCAAGTGGCTCAGGCAAGCACTCATGTGGCGATAGGTTT	5854
Qy	5331	caagccaagccagatggtgaagccgcgcttgccataggcagacaaaaccccaagcagcaa	5390
Db	5855	CCAGGCCAAGGCAGATGTTGTAAGCCGCGTTGCCATAGGCACACAAACCCCAAGCAGCAA	5914
Qy	5391	ccaatccatgcgaatcgtgtgataacgcacaagccaagggcgataatcattcgcatcgg	5450
Db	5915	CCAATCAATGGCCATCGGTGTATACGCNAAGCAAGCAAGCGGCGATCAATCATTCGCATCGG	5974

[illegible]

## RESULT

RESOLUTION 2  
US-08-968-685A-9  
Sequence 9, Application US/08968685A  
Patent No. 6214981  
GENERAL INFORMATION:  
APPLICANT: TUCKER, KENNETH  
APPLICANT: PLOSLA, LAURA  
TITLE OF INVENTION: MORAXELIA CAT  
PROTEIN-106 P  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Amer  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/968,685A  
FILING DATE: No. 6214981ember 12, 1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7969-060  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9542 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-968-685A-9

Query Match 65.0%; Score 4001.4; DB 4; Length 9542;  
Best Local Similarity 79.7%; Pred. No. 0;  
Matches 5115; Conservative 0; Mismatches 956; Indels 306; Gaps 21;

Qy 1 atgaatcacatctataaagtcatctttaaaagccacagggcacatttatggccgtggcg 60  
Db 218 ATGAATCACATCTATAAGTCAATCTTTAAACAAAGCCACAGGCACATTTATGGCGGTGGCG 277

Qy 61 gaatatgccaaatccccacagtcacgggggggtagctgtctacagggcaagtggcagt 120  
Db 278 GAATATGCCAAATCCCACAGCAGCGGGGGGGTAGCTGTCTACAGGCCAAGTTGGCAGT 337

Qy 121 gtacgcacttaagctttggccgtattgcccgcgtctcgctgcctcgtgagtcggtgcagc 180  
Db 338 GTACGCACCTCTGAGCTTTGCCCGATTGGCCGCTCGCTGCTCTCGTATCGGTGGCAGC 397

Qy 181 ctaaatggcagtgcttatgctcaacaaattactaccagaatgcgaattggtcacaacaaac 240  
Db 398 CTCAATGGCAGTGTATG-----CAGGTATCGGAATTAGTGAAGCAGAC 442

Qy 241 aagataaacaacagctgaaagcgatgcccctagcagagtgagcagatccatgctttt 300  
Db 443 GGGGA-----AAAGGGGAGCCAAATGCGCGGTGATTAATCCNTTGTATT 490

Qy 301 ggtagtcttttaaggcacaaaggctctcaagctattgtctatcggttagtgtaaacacagat 360  
Db 491 GGTGATATTGCTCAGGCACCTTGGCTCTCAATCTATTGCTATCGGTGACAAACAATAAGTT 550

Qy 361 cctaataatgtagtaaatgtaagtccacaggtccacaggaagtaacagagtcacagcc 420  
Db 551 CATAAATTCAAATAATAATGCTTAATATAGGTGCCAAAGCCTCAGGTAATGAGTCCATCGCC 610

Qy 421 atcgggtgtagtattggctgagggtagtcctcgtatgcatcggttagtgtagactta 480  
Db 611 ATCGGTGGTGTATGTTGGCTCTGGCTATGCCCTCGATTGCCATCGGTAGTGTAGTACGCTA 670

Qy 481 tatttgctaagaatcttgatctgaaga---atgaatttcaaaacttattcattggccat 537  
Db 671 TATTGAAAAGGAAACGGTACAGCAATCTCAGAGCTTCTACCTATTATTTCGCGGACAG 730

Qy 538 gaaatattaaaaaaatacaaacctcaaccgatggttaaaat---caaatatcgacgcaca 594  
Db 731 AAAGCATTTAAACGATATATACCAACTAGCTGACACTAATCTTCAAAAATATATAGACGCACA 790

Qy 595 agagcaaggcagccagctacgagtgagggagccatgtcatatgcacaggttcatttt 654  
Db 791 CAGGCACAGGACAGCCGCTACTGCTGAGTGGAGCCCATGTCTATATGCAAAAGGGTCATTTT 850

Qy 655 tccaacgcttttgtaatacagcaacagctgaagctgctctattccttggtgagtaggtctt 714  
Db 851 TCCAAGCGCTTTGGTACACGGGCAACAGCTGAAGTACCTATTCTTGGCAGTGGGTCTT 910

Qy 715 gccgcccaagccacaaaacaaatcttcaatcgctgtgtgttccaaatgcaaaagctaacgcg 774  
Db 911 ACCGCCACAGCCAAAGCAGACATCTTCAATCGCTGTGCTTAAATGCAACAAGCTATCGG 970

Qy 775 ttgcagcagacagccattggtggaataactgttagtattgggtcgagggcgttgcccta 834  
Db 971 TTGCAGGCACAGCCGTTGGTGGAGTACTCAAGTTAAATTTGAATCGAGGTATTGCCCTA 1030

Qy 835 ggttttggtcttcagatccttgatagggataataataacagatgcc-----gtgcc 885  
Db 1031 GGTTTTGGTTCTCAGGTCCTTCAGAGGATATGATGTAATGTCAGCAAAATGCTACGGCC 1090

Qy 886 tatgtaccactaggtataaaccttagcagacca---gtlaaagccaccgcgcaggggtgat 942  
Db 1091 TATGCACAGATGATAACACAGCCATACACCGGTATAAAGCCACCTTCAAGAAATGGT 1150

Qy 943 tctacgatataatttccattggaataataataataataataataataataataataataata 1002  
Db 1151 GCTACGGATGATTTTCCATTGGTANTAGTATGGGAATGA---CAGTATCAGGCGTAAA 1207

Qy 1003 atcatcaatgctggtggtggttctcgggataccgtagcggtcaatgtggcacagcttaaa 1062  
Db 1208 ATCATCAATGTCGCTGCGAGGTTCTCGGATACCGGATCGGTCATGTGCGACAGCTTAAA 1267

Qy 1063 ttggtgaggaactggtcgaataactctttaaagggtgagtggtgacaataatagc 1122  
Db 1268 GAGCGGTGAGGCTGGCTAATCGTCAATTTACTTTTAAAGGTGATGATAGCAAT----- 1321

Qy 1123 aatagctagaaaaggtttggaataactttaactttaagggtgagtcagcagcacaac 1182  
Db 1322 ANTAGTAGAAAAGGTTTGGCAAGACTTTAATCATCACAGGTGGTGCACAGACCAGC 1381

Qy 1183 gcattaaaccgaagctaacatcggtggtgtaaacagatggcaatggtcctgaagttaaactt 1242  
Db 1382 GCATTAAACCGATCATACATCGGTGCTACAAATGCGGATGGTCTCAAGTTCAACTT 1441

Qy 1243 gctaaagagctgactggtatgaccagtgctccgcacacaaacaaacacacgttagtaat 1302  
Db 1442 GCTGAAACTTTAAACAGAGCTTTAAATGGTTACCACCTGAAAACCTTAACGCCCAACGAGA 1501

Qy 1303 accaacaacaacagcgccgagcta----- 1326  
Db 1502 GTTACCGTAGCAAAACCCTTACCACAGATAAAATTTGTTTACCATGATATGAAT 1561

Qy 1327 ----- 1326

Db 1562 GGCAATTGATGAAAGCAAAACCTTATCTTGATAAAGACACTGGCATTCATGCAGGTGGTCAA 1621

Qy 1327 ----- 1326

Db 1622 AAGATTACCAAACTTACTGCTGGTGTAGTAGATGACGATGCGGCAACTTATGGACAGCTT 1681

Qy 1327 -----caagcggtggtttgaccttttagcccaataacacaggtaca 1365  
Db 1682 AAAAAAGTTAACCAAAACCGCTGAAAGTCTCTACAAACCTTTACCCTTTAAAAGGTAGAT 1741

Qy 1366 aaacagataaaacgcgtctacagcattgattgattgaagtttactaatgtagtaataagt 1425  
Db 1742 AAAAAATGTAATGATGCTAATGACAGCAAAATCATCACCGTGGGTAAAAATAAACAACCA 1801

Qy 1426 atagcaactaaggtaactactcgttatcccaaaaaaagaataatggttttctgctggtactaat 1485  
Db 1802 CAGGTACTCAAGTCAACACCCCTAAACCTTAAAGGTGAAAACGGTGTGATGTTTACAACC 1861

Qy 1486 gatgggtgtagaaagcaaaccttatcttcacaacagaaa---gctaaaagtggcaac 1542  
Db 1862 GAACAAATGTTACAGTTACCTTTGGGCTTTAACCAAAAATACGGTCTGACCGTTGGCAAC 1921

Qy 1543 agcacccctaaacagtggtgactgtactgttaataacacacactggttaataaacaacataccaa 1602  
Db 1922 AGCACCTTAAACACAGGATGGCTTATCTGTAAAAACACCACTAGTAAACAACAATCCAA 1981

Qy 1603 gtcggtgtaatggcattaaatttgccacagtcgcgtaataatgttgcacaatcacctcagca 1662







```

Db 6350 AACGAAATAAAGCCAAATGAGGATTTTCATCAGCGATGGGATGCGCTCCATGCCCAA 6409
Qy 5983 qccatcatctcgcagatccatggttaccgggggtattgccaccacacacgggtcaagg 6042
Db 6410 GCTACATCTCTGGCAGATCCATGTTACCGGGGTATTGCCACCCACACAGGTCAGAGT 6469
Qy 6043 gcggtgagcagtgagcagtgatgagtgatgagtgatgagtgatgagtgatgagtgat 6102
Db 6470 GCGTGGCAGTGGGACTGTGCAAGCTGTGCGATATGAGTGGTCAATGGGTATTTAAATCAAT 6529
Qy 6103 ggttcagcagatcccaagcagcagtgagggcgagtggtgagtggttttcaactt 6159
Db 6530 GGTTCAGCCGATACCCCAAGCCCATGTAGGGCGGCGAGTGGTGGTGCAGGTTTTCACATT 6586

RESULT 3
US-08-409-995-1
; Sequence 1, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 398-3249
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3294 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-409-995-1

Query Match 1.4%; Score 84.4; DB 1; Length 3294;
Best Local Similarity 50.0%; Pred. No. 8.3e-13;
Matches 245; Conservative 0; Mismatches 236; Indels 9; Gaps 1;

Qy 5677 aaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5736
Db 2803 AAAGCTGACGGTACTGCGGATATAAACCAAGGCGAAGTGAATGATAAGTTTCTACC 2862
Qy 5737 ttgtgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5796
Db 2863 GATGAAAACACAGTGTGACGCTTGATCCAAATGATCAATCAAAAGGTAAAGTGTCGTG 2922
Qy 5797 atccaaatgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5856
Db 2923 ATTGCAATGTGGCTAATGGCGATATTTCTCCACTTCCACCGATGCGATTAAACGGAAGT 2982

```

```

Qy 5857 cagttgtac-----aaagccaccacaaagcattgcacacgcaacacaaatgagcgttgcac 5907
Db 2983 CAGTTGTATGCTGTGTCACAAAGGGGTAAACAACCTTGTCTGGCAAGTGAATAAATCTTGAG 3042
Qy 5908 catcgtatccacacacacacacacacacacacacacacacacacacacacacacacacacac 5967
Db 3043 GCGAAAGTGAATAAAGTGGGCAACGTCAGATGCGAGGTACAGCAAGTGCATTAGCGGCT 3102
Qy 5968 gctccatcagcacaacacacacacacacacacacacacacacacacacacacacacacacacac 6027
Db 3103 TCACAGTTACCAACAGCCACTATGCCAGGTAAATGTTGCTATTCCGGGAAGTACT 3162
Qy 6028 cacaacggtcaaggtcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 6087
Db 3163 TATCAAGGTCAAAATGTTTAGCTATCGGGGTATCAAGATTTCCGATATATGCAAAAGTG 3222
Qy 6088 gtatttaaatcaatggttcagcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6147
Db 3223 ATTATTGCTGTGTCAGGCACACCAATAGTCAAGTAAACAGGCGTTGCGAGCGGTGT 3282
Qy 6148 ggttttcaact 6157
Db 3283 GGTACCAGT 3292

RESULT 4
US-08-685-467-1
; Sequence 1, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 398-3249
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3294 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-685-467-1

Query Match 1.4%; Score 84.4; DB 3; Length 3294;

```



TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP  
 STREET: Four Embarcadero Center, Suite 3400  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States  
 ZIP: 94111-4187  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/913,942  
 FILING DATE: 29-DEC-1997  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/409,995  
 FILING DATE: 24-MAR-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US96/4031  
 FILING DATE: 22-MAR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Vargoe, Dolly A.  
 REGISTRATION NUMBER: 39,054  
 REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 781-1989  
 TELEFAX: (415) 398-3249  
 TELEX: 910 277299  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7291 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: unknown  
 TOPOLOGY: unknown  
 MOLECULE TYPE: DNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 163..7221  
 PS-08-913-942-3

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,942
FILING DATE: 29-DEC-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA: PCT/US96/4031
APPLICATION NUMBER: 22-MAR-1996
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vance, Dolly A.
REGISTRATION NUMBER: 39,054
REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 2037 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-913-942-14

Query Match 1.48; Score 83.6; DB 4; Length 2037;
Best Local Similarity 54.1%; Pred. NO. 1e-12;
Matches 198; Conservative 0; Mismatches 159; Indels 9; Gaps 1;

QY 5801 aaaaatgtgcagcaggtgaggtcagtcgcccacgacccgcatcggtgtaatggttagcaggt 5860
Db 1670 AAAAACTCCGCGCAGCGCGAAATTTTCGCACCTCCACCGATGCGATTAAACGGAAGCCAGT 1729

QY 5861 tgtac-----aaagccaccacaaagcattgccaacgcaacaaatgagcttgaccalc 5911
Db 1730 TGTATGCCGTGCGCAAAAAGGGGTAAACAAACCTTGTGGAACAAGTGAATAATCTTGAGGGCA 1789

QY 5912 gtatcacacaaacgaataaagccaatgcagggatttcacgcgatggcgatggcgt 5971
Db 1790 AAGTGAATAAAGTGGCAAAACGTCGAGATGCAAGGTACTGCAAGTGCATTAGCGCCTTAC 1849

QY 5972 ccattgcacaagctctacattctctgcatcatcattgttaccgggggtattgcacaccaca 6031
Db 1850 AGTTACCACAAGCCACTATGCCAGGTAAATCAATGGTTTCTATTTCGCGGAAGTAGTTATC 1909

QY 6032 acggttcgaagtgccgggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggt 6091
Db 1910 AAGGTCAAAATGGTTTATGCTATCGGGGTATCAAGAAATTTCCGATTAATGCGCAAGTGATTA 1969

QY 6092 ttaaaatcaatgttcagcgcatacccaaggccatgttagggcgccaggttggtgcaggtt 6151
Db 1970 TTCGCTGTCTGCGCAACCAATATCAAGGTAAACAGCGCTTGACAGCAGGTCTGTGTT 2029

QY 6152 ttact 6157
Db 2030 ACCAGT 2035

```

```

RESULT 7
US-08-913-942-3
; Sequence 3, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph
; APPLICANT: Barenkamp, Stephen J

```

Db	2558	CTTTAACT---	GAGCCAAAGCGGGGAGCAAGCTTTCACACGTTGATTTAAATGTGGATG	2614
Qy	2483	aaaaagacaaatccaa	cgctgcagcattgatgtgctaaatgcaggtcttaacctaa	2542
Db	2615	CGAGGAAAATCCAT	GAGCAAGTATTGAGATGTATTGCGCGCAGTTGGATATTC	2674
Qy	2543	aaaataatggcaaa	gaaagactttgtctccacttatgcactgttgactttatcgatg	2602
Db	2675	AAGTAATGGTAATA	TGTTGATTATGTAGCGACGTATGACACAGTAAACTTTTACCGATG	2734
Qy	2603	gcaatgccaccaccg	ccacagtaacttatgatgaagccaatacaaacacagtaaaagtggcgt	2662
Db	2735	ACAGCACAGGTGA	CAACAACGGTAAC-----CGTAACCCAAAAGACAGATGGCAAGGTG	2788
Qy	2663	atgatgtgaatgg	atgatgagaaccattgaactgacaggcgataatggcaa	2714
Db	2789	CTGACGTTAAAT	CGTGGCAAAACTTCTGTTATCAAGACCAACCAACGGCAA	2840
RESULT 9				
US-08-685-467-3				
: Sequence 3, Application US/08685467				
: Patent No. 6060059				
: GENERAL INFORMATION:				
: APPLICANT: St. Gene III, Joseph W.				
: APPLICANT: Barenkamp, Stephen J.				
: TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS				
: NUMBER OF SEQUENCES: 6				
: CORRESPONDENCE ADDRESS:				
: ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert				
: STREET: Four Embarcadero Center, Suite 3400				
: CITY: San Francisco				
: STATE: California				
: COUNTRY: United States				
: ZIP: 94111-4187				
: COMPUTER READABLE FORM:				
: MEDIUM TYPE: Floppy disk				
: COMPUTER: IBM PC compatible				
: OPERATING SYSTEM: PC-DOS/MS-DOS				
: SOFTWARE: Patent In Release #1.0, Version #1.30				
: CURRENT APPLICATION DATA:				
: APPLICATION NUMBER: US/08/685,467				
: FILING DATE: 22-JUL-1996				
: CLASSIFICATION: 424				
: PRIOR APPLICATION DATA:				
: APPLICATION NUMBER: US 08/409,995				
: FILING DATE: 24-MAR-1995				
: ATTORNEY/AGENT INFORMATION:				
: NAME: Silva, Robin M.				
: REGISTRATION NUMBER: 38,304				
: REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS				
: TELECOMMUNICATION INFORMATION:				
: TELEPHONE: (415) 781-1989				
: TELEFAX: (415) 398-3249				
: TELEX: 910 277299				
: INFORMATION FOR SEQ ID NO: 3:				
: SEQUENCE CHARACTERISTICS:				
: LENGTH: 5736 base pairs				
: TYPE: nucleic acid				
: STRANDEDNESS: unknown				
: TOPOLOGY: unknown				
: MOLECULE TYPE: DNA (genomic)				
US-08-685-467-3				
Query Match 1.1%; Score 66; DB 3; Length 5738;				
Best Local Similarity 49.4%; Pred. No. 1,1e-07;				
Matches 233; Conservative 0; Mismatches 230; Indels 9; Gaps				
Qy	2243	ttaattacgaccaata	gagccacaggttacagttacctttggcattgacccaagaatg	2302
Db	2378	TTGAGGTGAAAC	TGCGAAAGTGAGTGACTTAAACGATTGGCGGGGAATACACCTACAG	2437







Thu Sep 13 14:19:08 2001

```
RESULT 15
US-09-377-155-20
; Sequence 20, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 1776
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1773)
US-09-377-155-20

Query Match      0.9%   Score 55.2; DB 4; Length 1776;
Best Local Similarity 51.2%; Pred No. 4.7e-05;
Matches 129; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 5906 accatcgtatccacacaaacgaaataaagccaatgcaggatttcacgcgatggcga 5965
   || || || || || || || || || || || || || || || || || || || ||
DB 1520 acaaccgatcgacaatgtggacgcacgcgctgcggcgatcgcccaagcgattgcaa 1579
   || || || || || || || || || || || || || || || || || || || ||

QY 5966 tggcgctccatgccacaagcctacattcctgcgagatccatggttacccggggtattgcc 6025
   || || || || || || || || || || || || || || || || || || || ||
DB 1580 ccgcagggtctggttcaggcgtatttgccggcgaagagatgatgcgatcggcgcggca 1639
   || || || || || || || || || || || || || || || || || || || ||

QY 6026 ccacacaagggtcaagtgcggtggcagtgggactgtggaagctgtcggaataatggtcaat 6085
   || || || || || || || || || || || || || || || || || || || ||
DB 1640 cttatcgcggcgaagcgggttacgscatcggtactccagtatttccgacggcgaaatt 1699
   || || || || || || || || || || || || || || || || || || || ||

QY 6086 ggggtatttaaaatcaatgttcagccgatacccaagcccatgtagggcggcagttggtg 6145
   || || || || || || || || || || || || || || || || || || || ||
DB 1700 ggattatcaaaaggcagcgtctccggcaattccgcgcaatttcgggtgcttccgcacatcg 1759
   || || || || || || || || || || || || || || || || || || || ||

QY 6146 cagggttttcaact 6157
   || || || || || || || || || || || || || || || || || || || ||
DB 1760 tcgggttatcagt 1771
   || || || || || || || || || || || || || || || || || || || ||
```

Search completed: September 12, 2001, 12:57:17  
Job time: 1059 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2001, 03:34:16 ; Search time 18125.9 Seconds  
(without alignments)  
3211.985 Million cell updates/sec

Title: us-09-361-619-8  
Perfect score: 6159  
Sequence: 1 atgaatcacatcataaagt.....ttggtgcagggttttcacttt 6159

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_est3.\*
- 4: gb\_est4.\*
- 5: gb\_est5.\*
- 6: gb\_est6.\*
- 7: gb\_est7.\*
- 8: gb\_est8.\*
- 9: gb\_est9.\*
- 10: gb\_est10.\*
- 11: gb\_est11.\*
- 12: gb\_est12.\*
- 13: gb\_est13.\*
- 14: gb\_est14.\*
- 15: gb\_est15.\*
- 16: gb\_est16.\*
- 17: gb\_est17.\*
- 18: gb\_est18.\*
- 19: gb\_est19.\*
- 20: gb\_est20.\*
- 21: gb\_est21.\*
- 22: gb\_est22.\*
- 23: gb\_est23.\*
- 24: gb\_est24.\*
- 25: gb\_est25.\*
- 26: gb\_est26.\*
- 27: gb\_est27.\*
- 28: gb\_est28.\*
- 29: gb\_est29.\*
- 30: gb\_est30.\*
- 31: gb\_est31.\*
- 32: gb\_est32.\*
- 33: gb\_est33.\*
- 34: gb\_est34.\*
- 35: gb\_est35.\*
- 36: gb\_est36.\*
- 37: gb\_est37.\*
- 38: gb\_est38.\*
- 39: gb\_est39.\*
- 40: gb\_est40.\*
- 41: gb\_est41.\*
- 42: gb\_est42.\*
- 43: gb\_est43.\*
- 44: gb\_est44.\*
- 45: gb\_est45.\*
- 46: gb\_est46.\*
- 47: gb\_est47.\*

- 44: em\_esthum10.\*
- 45: em\_esthum11.\*
- 46: em\_esthum12.\*
- 47: em\_esthum13.\*
- 48: em\_esthum14.\*
- 49: em\_esthum15.\*
- 50: em\_esthum16.\*
- 51: em\_esthum17.\*
- 52: em\_esthum18.\*
- 53: em\_esthum19.\*
- 54: em\_esthum20.\*
- 55: em\_esthum21.\*
- 56: em\_esthum22.\*
- 57: em\_esthum23.\*
- 58: em\_esthum24.\*
- 59: em\_esthum25.\*
- 60: em\_esthum26.\*
- 61: em\_esthum27.\*
- 62: em\_esthum28.\*
- 63: em\_estin1.\*
- 64: em\_estin2.\*
- 65: em\_estin3.\*
- 66: em\_estin4.\*
- 67: em\_estin5.\*
- 68: em\_estom1.\*
- 69: em\_estom2.\*
- 70: em\_estov1.\*
- 71: em\_estov2.\*
- 72: em\_estpl1.\*
- 73: em\_estpl2.\*
- 74: em\_estpl3.\*
- 75: em\_estpl4.\*
- 76: em\_estpl5.\*
- 77: em\_estpl6.\*
- 78: em\_estpl7.\*
- 79: em\_estpl8.\*
- 80: em\_estpl9.\*
- 81: em\_estpl10.\*
- 82: em\_estro1.\*
- 83: em\_estro2.\*
- 84: em\_estro3.\*
- 85: em\_estro4.\*
- 86: em\_estro5.\*
- 87: em\_estro6.\*
- 88: em\_estro7.\*
- 89: em\_estro8.\*
- 90: em\_estro9.\*
- 91: em\_estro10.\*
- 92: em\_estro11.\*
- 93: em\_estro12.\*
- 94: em\_estro13.\*
- 95: em\_estro14.\*
- 96: em\_estro15.\*
- 97: em\_estro16.\*
- 98: em\_estro17.\*
- 99: em\_estro18.\*
- 100: em\_estro19.\*
- 101: em\_estro20.\*
- 102: gb\_est25.\*
- 103: gb\_est26.\*
- 104: gb\_est27.\*
- 105: gb\_est28.\*
- 106: gb\_est29.\*
- 107: gb\_est30.\*
- 108: gb\_est31.\*
- 109: gb\_est32.\*
- 110: gb\_est41.\*
- 111: gb\_est42.\*
- 112: gb\_est43.\*
- 113: gb\_est44.\*
- 114: gb\_est45.\*
- 115: gb\_est46.\*
- 116: gb\_est47.\*

us-09-361-619-8.rst

Thu Sep 13 14:19:16 2001

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



Thu Sep 13 14:19:16 2001

```

Db 255 TAATGATATCATAGTAATAATAATAATAATAATAATAATGCTAATAGTAACGCTATATA 196
QY 2007 cagtaatggcaccagtggttaataataa 2033
Db 195 TGATAGTAGTGACAAATCATATAACACTGA 169

RESULT 2
LOCUS AQ946120 641 bp DNA GSS 27-JAN-2000
DEFINITION Sheared DNA-46J23.TF Sheared DNA Trypanosoma brucei genomic clone
ACCESSION AQ946120
VERSION AQ946120.1 GI:6769385
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
REFERENCE 1 (bases 1 to 641)
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Doneison,J.,
Fraser,C. and Adams,M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: Sheared DNA-46J23.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
http://www.tigr.org/tdb/mbd/tbdb/.
Seq primer: M13-Reverse
Class: shotgun.
Location/Qualifiers
1. 641
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-46J23"
/clone_lib="Sheared DNA"
/notes="Vector: pUC18; Site_1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Barell, Oxford University
Press, 1999)."

BASE COUNT 213 a 35 c 10 g 383 t
ORIGIN

Query Match 1.3%; Score 77.4; DB 236; Length 641;
Best Local Similarity 46.3%; Pred. No. 1.9e-09;
Matches 255; Conservative 0; Mismatches 296; Indels 0; Gaps 0;

QY 1555 agtggtagctgactgtaataacaccctggtaataacaaatccagtcggtgcta 1614
Db 638 ACTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 579
QY 1615 ggcattaaattggccacagtcgctataatgttgcaaatccctcagcaacagtcggcact 1674
Db 578 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 519

```

```

QY 1675 gctcgattaccgaagagaaatlggttttctggtactactaatgatggagttgatgaaca 1734
Db 518 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 459
QY 1735 gcaccatatttggataaagaacgacttaagtg99gctggttgaaattaccacagatagt 1794
Db 458 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 399
QY 1795 ggtattaatgctggttaacacagattaccggacttactaaatgggtatagcaaatcccgat 1854
Db 398 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 339
QY 1855 gcggttaccatcaaacagctcaagacgcgaagcctactttaaacccagcgatggcaltc 1914
Db 338 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 279
QY 1915 agtattaataglaataacggggatctctagttgtagtagtggaatattaccaccccaact 1974
Db 278 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 219
QY 1975 tatacattagcgtgaacacacacacacacacacacacacacacacacacacacacacacac 2034
Db 218 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 159
QY 2035 tttagttagttagttagttagttagttagttagttagttagttagttagttagttagttagt 2094
Db 158 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 99
QY 2095 tatctaaataa 2105
Db 98 GATAATAATAA 88

RESULT 3
LOCUS AQ940248 700 bp DNA GSS 27-JAN-2000
DEFINITION Sheared DNA-42E21.TF Sheared DNA Trypanosoma brucei genomic clone
ACCESSION AQ940248
VERSION AQ940248.1 GI:6763513
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
REFERENCE 1 (bases 1 to 700)
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Doneison,J.,
Fraser,C. and Adams,M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: Sheared DNA-42E21.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
http://www.tigr.org/tdb/mbd/tbdb/.
Seq primer: M13-Forward
Class: shotgun.
Location/Qualifiers
1. 700
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-42E21"
/clone_lib="Sheared DNA"

FEATURES
source

```





```

AU086536      500 bp      mRNA      EST      27-JAN-2001
LOCUS          Sugano Malaria cDNA library Plasmodium falciparum CDNA
DEFINITION     clone XPFn2783, mRNA sequence.
ACCESSION      AU086536
VERSION        AU086536.1  GI:12388677
KEYWORDS       malaria parasite P. falciparum.
SOURCE         Plasmodium falciparum
ORGANISM       Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE      1 (bases 1 to 500)
AUTHORS        Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
TITLE          FULL-malaria: a database for a full-length enriched cDNA library
               from human malaria parasite, Plasmodium falciparum
JOURNAL        Nucleic Acids Res. 29 (1), 70-71 (2001)
MEDLINE        20574754
COMMENT        Contact: Junichi Watanabe
               Institute of Medical Science
               The University of Tokyo, Department of Parasitology
               4-6-3, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
               Tel: 81-3-5449-5378
               Fax: 81-3-5449-5410
               Email: jwatanabe@nagane.ims.u-tokyo.ac.jp
               Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano
               , S. Construction and characterization of a full length-enriched and
               a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES       Location/Qualifiers
               1..500
               /organism="Plasmodium falciparum"
               /db_xref="3D7"
               /db_xref="taxon:5833"
               /clone="XPFn2783"
               /clone_lib="Sugano Malaria cDNA library"
FEATURES       source

```

BASE COUNT	278 a	43 c	34 g	142 t	3 others
ORIGIN					
Query Match	1.1%	Score 66.8;	DB 107;	Length 500;	
Best Local Similarity	47.9%;	Pred. No. 1.2e-06;			
Matches 191;	Conservative	0;	Mismatches 208;	Indels 0;	Gaps 0;
Qy 1714	aatgatgagtgatgagaacaagcaccata	tttggataaagaacagcacttaagctggggctgt	1773		
Db 86	ATGATAATAATATGCAANTATTACCGTATGCAANTARTAT	TGGAANTARTATGGAANTARTANATATGAA	145		
Qy 1774	gttgaataattaccacagatagtggtattaatgtctggt	taatacacaagattaccggacttact	1833		
Db 146	AATAAAAAATTAATGAATAATAATAATAATAAACAA	CAATAATAATAAACAAATAATAAT	205		
Qy 1834	aatggtatagcaaataccagatgcgggtaccatacaaac	agctcgaagacgcgaagccctact	1893		
Db 206	AATAATAATAACANATAATAATAATAATAATAATAA	TAAATAATAATAATAATAATAATAATAATAATAAT	265		
Qy 1894	ttaaacgcagggcgatggcctcagttaatatagtaata	aacgggctcagttgtagtagt	1953		
Db 266	AGCAACAATAATATATAGTAACAACAATAATAATAA	TAGTAACAACAATAATAATAATAATAATAATAAT	325		
Qy 1954	ggcaatattaccaccccccaattataacattagctg	tgagaacccactaagcttaacagtaat	2013		
Db 326	AATAATAATAATAATAGTAACAATAATAATAATAA	TATAGTAACAACAATAATAATAATAATAATAATAC	385		
Qy 2014	ggcacaggtgtaataataaaattaggttagtagtaat	gctcatgatacaaatagcttagtt	2073		
Db 386	AANTATACANATAANTATATAATAGTACACATATA	TATTACANATAATAGTAAACAATAAT	445		
Qy 2074	accggcaagaagatttggcgagactactctaaataa	gaagctaat	2112		
Db 446	AATAATATCATAGTACCCCTAATTAGTAAATAGTAAT	AGTAAT	484		

RESULT  
AQ941683

LOCUS	AQ941693	621 bp	DNA	GSS	27-JAN-2000
DEFINITION	Sheared DNA-43D2.TR Sheared DNA Trypanosoma brucei genomic clone				
ACCESSION	AQ941683	Sheared DNA-43D2, DNA sequence.			
VERSION	AQ941683.1	GI:6764948			
KEYWORDS	GSS.				
SOURCE	Trypanosoma brucei.				
ORGANISM	Trypanosoma brucei				
	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.				
REFERENCE	1 (bases 1 to 621)				
AUTHORS	El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M.				
TITLE	Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library				
JOURNAL	Unpublished (1999)				
COMMENT	Other_GSSs: Sheared DNA-43D2.TF Contact: Najib M. El-Sayed Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: meisayed@tigr.org Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: <a href="http://www.tigr.org/tdb/mdb/tbdb/">http://www.tigr.org/tdb/mdb/tbdb/</a> . Seq primer: M13-Reverse Class: shotgun.				
FEATURES	Location/Qualifiers				
source	1. 621 /organism="Trypanosoma brucei" /strain="TREU927/4 GUTat 10.1" /db_xref="taxon:5691" /clone="Sheared DNA-43D2" /clone_lib="Sheared DNA" /note="Vector: pUC18; Site.1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."				
BASE COUNT	318 a	15 c	78 g	210 t	
ORIGIN					

[illegible]

```

Qy 1988 tgaacacactaagcttaacagtagtgccaccagtggtgtaataataaaatttagttagta 2047
Db 512 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 571

Qy 2048 atgctcatgatacaat 2064
Db 572 ATAATAATAATAATAAT 588

RESULT 7
LOCUS AU076358 469 bp mRNA EST 27-JUL-1999
DEFINITION AU076358 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoideum cDNA clone SSA642, mRNA sequence.
ACCESSION AU076358
VERSION AU076358.1 GI:5607256
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
AUTHORS 1 (bases 1 to 469)
TITLE Dictyostelium discoideum.
JOURNAL Urushihara, H.
COMMENT Developmental cDNA in Dictyostelium discoideum (1999)
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@sakura.cc.tsukuba.ac.jp
PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES
source
1. 469
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSA642"
/clone_lbb="dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug" 26 g 157 t 3 others
BASE COUNT 253 a 30 c 157 t
ORIGIN

Query Match 1.08; Score 62.4; DB 107; Length 469;
Best Local Similarity 50.08; Pred. No. 1.8e-05;
Matches 153; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

Qy 1776 tgaattaccacagatggtgtattaatgctggttaacacagattaccgagctactaa 1835
Db 38 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 97

Qy 1836 tggtagcgaataaccgtagcggttaccatcaaacagctcaaacgagcgcctacttt 1895
Db 98 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 157

Qy 1896 aaacgagcgatggcagcatgataatagtaataaacggggatctagttgtagtagtg 1955
Db 158 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 217

Qy 1956 caattattaccaccccaacttaacattagcggtgaaacacacacagcttaacagta 2015
Db 218 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 277

Qy 2016 caccagtggttaataaatttagttagtgcctcatgatacaacatgcttagttac 2075
Db 278 TAATAATGATAATGATAATGATAATGATAATGATAATGATAATGATAATGATAATA 337

Qy 2076 cgccaa 2081
Db 338 TAGCAA 343

```

RESULT 8  
CNS04KUS/c

```

LOCUS CNS04KUS 541 bp DNA GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
117023 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL295309.1 GI:8033889
VERSION AL295309.1
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Acanthomorphi; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 541)
AUTHORS Roest-Crolius, H., Jaillon, O., Dasilva, C., Fzames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 541)
AUTHORS Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fzames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 541)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.
FEATURES
source
1. 541
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="117023"
/clone_lbb="G"
/note="Genoscope sequence ID : COBG117AH12LP1-end : T7"
BASE COUNT 152 a 88 c 50 g 214 t 37 others
ORIGIN

Query Match 1.08; Score 62; DB 221; Length 541;
Best Local Similarity 44.98; Pred. No. 2.5e-05;
Matches 124; Conservative 26; Mismatches 126; Indels 0; Gaps 0;

Qy 1790 atagtggtattaatgctgtaacacagattaccgagctactactaatggtatagcaata 1849
Db 511 ATAATAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 452

Qy 1850 ccgtagcggttaccatcaaacagctcaaacgagcgcctactttaaaccgagcgatg 1909
Db 451 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 392

Qy 1910 gcatcagtaataatgtaaacgggggtagctagttgtagtagtgccatattaccacc 1969
Db 391 ATAATAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 332

Qy 1970 caactataacattagctgaaacacacacacacacacacacacacacacacacacac 2029
Db 331 ATATMATMATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 272

Qy 2030 ataaatttagttagttagttagttagttagttagttagttagttagttagttagttag 2065
Db 271 ATAATAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 236

```

RESULT 9  
CNS02156

[illegible]





Search completed: September 13, 2001, 03:34:48  
Job time: 53710 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:37:02 ; Search time 78.73 Seconds  
(without alignments)  
1580.859 Million cell updates/sec

Title: US-09-361-619-9  
Perfect score: 10356  
Sequence: 1 MNHKKYFNKATGTGTFMAVA.....NGSADTQCHVCAAVGAGPHF 2053

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0601.\*  
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10356	100.0	2053	22 AAB69135	M. catarrhalis str
2	7675	74.1	2047	22 AAB69134	M. catarrhalis str
3	7408	71.5	1992	22 AAB69133	M. catarrhalis str
4	7405	71.5	1992	17 AA04505	Moraxella 200 kDa
5	7405	71.5	1992	22 AAB69137	M. catarrhalis M56
6	7219.5	69.7	2314	22 AAB69136	M. catarrhalis les
7	1206	11.6	2353	17 AAR9393	Haemophilus adhesi
8	1205	11.6	2411	21 AAB23860	Haemophilus influe
9	1021	9.9	1104	21 AAB23856	Haemophilus influe
10	1021	9.9	1104	21 AAB23859	Haemophilus influe
11	994	9.6	1004	21 AAB23857	Haemophilus influe

12	952	9.2	1002	21	AAB23854	Haemophilus influe
13	809	7.8	2039	19	AAW56322	Haemophilus paraga
14	779.5	7.5	2042	19	AAW56319	Haemophilus influe
15	710	6.9	1094	21	AAB23858	Haemophilus influe
16	690	6.7	1098	17	AAR99392	Haemophilus adhesi
17	583.5	5.6	679	17	AAR99394	Haemophilus adhesi
18	583.5	5.6	679	21	AAB23855	Haemophilus influe
19	553	5.3	2514	21	AAW75097	Neisseria meningit
20	549.5	5.3	1601	18	AAW30292	Non-typeable Haemo
21	526.5	5.1	2383	21	AAB15945	E. coli proliferat
22	513	5.0	1598	18	AAW30291	Non-typeable Haemo
23	512.5	4.9	2599	21	AAW75098	Neisseria meningit
24	512	4.9	1529	14	AAW41732	High molecular wei
25	506	4.9	2048	21	AAW75096	Neisseria gonorrh
26	484.5	4.7	2893	19	AAW98828	H. pylori GHP0 148
27	484.5	4.7	2893	19	AAW71556	Helicobacter polyp
28	468	4.5	1477	14	AAW41724	High molecular wei
29	462	4.5	1477	18	AAW30294	Non-typeable Haemo
30	459	4.4	1338	14	AAW41731	High molecular wei
31	458	4.4	1477	14	AAW41728	High molecular wei
32	457	4.4	1536	15	AAW63505	Haemophilus high m
33	457	4.4	1536	21	AAB01846	Haemophilus influe
34	456	4.4	1536	14	AAW41723	High molecular wei
35	456	4.4	1536	18	AAW30293	Non-typeable Haemo
36	454.5	4.4	1536	14	AAW41725	High molecular wei
37	452	4.4	1477	15	AAW63506	Haemophilus high m
38	451	4.4	1222	21	AAB01830	H. influenzae stra
39	451	4.4	1228	21	AAB01828	Haemophilus influe
40	451	4.4	1477	21	AAB01848	Haemophilus influe
41	451	4.4	1638	20	AAW00138	Enterococcus faeca
42	451	4.4	1638	20	AAW00140	Enterococcus faeca
43	451	4.4	1638	20	AAW00142	Enterococcus faeca
44	449	4.3	2902	22	AAB46351	H. pylori HPN165 p
45	440	4.2	1612	19	AAW65088	R. prowazekii S-la

## ALIGNMENTS

RESULT 1  
AAB69135  
ID AAB69135 standard; Protein; 2053 AA.  
XX AAB69135;  
XX  
XX 24-APR-2001 (first entry)  
XX  
XX M. catarrhalis strain Q8 200kDa protein SEQ ID NO:9.  
XX  
XX Moraxella catarrhalis strain Q8; major outer membrane protein;  
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection.  
XX  
XX Moraxella catarrhalis.  
XX  
XX WO200107619-A1.  
XX  
XX 01-FEB-2001.  
XX  
XX 26-JUL-2000; 2000WO-CA00870.  
XX  
XX 27-JUL-1999; 99US-0361619.  
XX  
XX (CONN-) CONNAUGHT LAB LTD.  
XX  
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;  
XX WPI; 2001-159722/16.  
XX N-PSDB; AAF59104.  
XX  
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
XX useful in protective vaccines and for diagnosis

Claim 1; Fig 4A-V; 247pp; English.

ps The present invention describes an isolated and purified nucleic acid (I) CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. CC The 200 kDa outer membrane protein (II) has antibacterial activity and CC can be used in vaccines. (II), and its truncated versions, are used as CC immunogenic compositions and vaccines to protect against M. catarrhalis CC infections, particularly otitis media in humans. (II) is also used as CC antigen in immunoassays for detecting specific antibodies (Ab), and to CC generate Ab. (I) are used for recombinant production of (II) and its CC fragments are used as probes for identifying/cloning 200 kDa protein CC genes from other strains, and for diagnostic detection of M. catarrhalis. CC (I) makes possible production of large amount of recombinant immunogens. CC Expression of truncated versions of (II) reduces toxicity of the protein CC towards the Escherichia coli host. The present sequence represents the CC M. catarrhalis strain Q8 200kDa protein, which is given in the CC exemplification of the present invention.

xx Sequence 2053 AA;

Query Match 100.0%; Score 10356; DB 22; Length 2053;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2053; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNHIVKINATGTPMAVEYAKSHSTGGSCATGQVGSVETLSFARIALAVLIVIGAT 60  
DB 1 mnhivkvinfatgtfmaevayakshstggscatgqvgsvetlsfariaalavlivigat 60  
QY 61 LNSAYAAQQTITTEIGQTNKINNTLKGDALATGEASIAFGSLSAQGSQAIAISVRPD 120  
DB 1 lnsayaqqitttiegqtnkinntlkgdalatgeasiafsgslsaqgsqaiaisvrpd 120  
QY 121 PNNGSNGNVSAGHAKNESIAIGDGVLAEGDASTAIGSDDLPLPKNLDLKNFPHKLHGHE 180  
DB 121 pnngsngnvsaghaknesiaigdvlaegdastaigsdldlplknlldlknfphklhghe 180  
QY 181 ILKKIQSTGKTKYRTRAGHASTAVGAMSAQGHFSNAGFTYATAPAAVSLAVGLAA 240  
DB 181 ilkkiqstgdktkyrtraghastavgamsaqghfsnagfityatapaaavslavglaa 240  
QY 241 QATKQSSIAVGSNAKANAFATAGTGNVNLGRGVALFGSGOILDRDNTDASAVVPLG 300  
DB 241 qatkqssiavgssnakanafatagtnvnlgrgvvalfgsgoildrdntdasavvplg 300  
QY 301 KTLADQVKATRQGDSTDFISIGNSNNSIRRKINVGAGSRDTPAVNVAQILKVEELA 360  
DB 301 ktladqvkatrqgdstdfisignsnnsirrkinvagssrdtpavnvaqilklveela 360  
QY 361 NRKITFKGDDNNSNSVERGLGNTLTIKGDAQTNALTEANIGVVTGNGLKKVLAKELTG 420  
DB 361 nrkitfkdgddnnsnsverglgntltikgdaqtnalteanigvvtgnglkkvlakeitg 420  
QY 421 LTSVSATNKIVNTNNNAELOSGLTFSPITGKTDTKTVISIDGLKFTNDSNSTATKG 480  
DB 421 ltsvsatnkivntnnnaelosgltfspitgktdtktvvisidglkftndsnstatk 480  
QY 481 TTRITKKIKFAGTNDGDESKPYLDNEKLVGNSTLNSGLFTVNTTGNKQIQVGANGI 540  
DB 481 ttritkkikfagtndgdeskpyldneklvgnstlnsglftvnttgnkqiqvgangi 540  
QY 541 KPATVANNVANTSATVGTARITEEKIFAGTNDGDEAPYLDKERLKVGRVEITTDGSI 600  
DB 541 kpatvannvantSATVGTARITEEKIFAGTNDGDEAPYLDKERLKVGRVEITTDGSI 600  
QY 601 NAGNHKITGLTNGIANTDAVITKOLDKAKPTINAGDGIISNSNNGDLVDSSGNTTPTYN 660  
DB 601 nagnhkitglTNGIANTDAVITKOLDKAKPTINAGDGIISNSNNGDLVDSSGNTTPTYN 660  
QY 661 ISVKTTLKNSGNTSGNNFVSNAHDNLSLVTAKDLADYLNKVNETAADSLPSEKVVQNGD 720  
DB 661 isvkttlknsngntsgnnfvsnahdnlslvtakdladylnkvnetaadslpsekvvqngd 720

QY 721 NSNNAITVCKDNGKNTENTLKLKGNGVNTTNRATGTVTFGIDQSNGLTTPKLVGSDT 780  
DB 721 nsnnaityvckdngkntentlklkgngvnttnratgtvtfgidqsnnglttppklvgsdt 780  
QY 781 NGNRLVIEQVPSADGNSNTNIIKGLSPTLPSTASPSGRNIALGNTIEKDKSNAASIDDV 840  
DB 781 ngnrlvieqvpSadgnsntniiKGLSptlpstaspsgrniaLgntieekdkSnaasiddv 840  
QY 841 LNAFNLKNGKDKDFSVSTVDVDFIDGNATTTATVYDEANQTSKVAYDVNVDKTIET 900  
DB 841 lnafnlnknngkdkdfsvstvdvdfidgnatttATVYdeanQTSkvayDVnvdktiet 900  
QY 901 GDNGKQKLGKTKIKLTETSTNGNATFTSTDDHALVKASDIAGNLNTLAEETHTKGTAN 960  
DB 901 gdngkkqlgvtikltetstngnatftstddhalvkasdiagnlntlaeethtkgtan 960  
QY 961 TALQTFVVKVVDENDKADDTNATVCGKDTSGKVNTLKLKGNGLDITDKDCTVTFGIN 1020  
DB 961 talqtfvkvvdendkaddtnatlvkgdtsGkvntllKlkgnglditdkdctvtfgin 1020  
QY 1021 TOSGLKAGDSTTLNNGLSIKNTASNEQIQVGADGVKVFAMVNGVVGAGIDGTRTRTRDE 1080  
DB 1021 tqsglkagdsttlnnglsikntasneqiqvgadgvkfamvngvvgagidgttrtrde 1080  
QY 1081 IGFTGTNGSLDKSKPHLSKDGINAGGKKITNIOSETAKNSHDAVTGGKTYDLKTELENK 1140  
DB 1081 igftgtngsldkskphlskdginagggkkitnIdsegeiaknshdavggklydlkteleNk 1140  
QY 1141 ISSPATAQNSLHRSFVADGQGNFTVSNPYSSYDTSKTSDDVITFAGENGITTKVNGGV 1200  
DB 1141 issstaktaqnsLhfsfvadeggnftvsnpysyDtsktsdsvltfagengittkvnggv 1200  
QY 1201 RVGIDQTKGLTPPKLTVGNNNGKGIIVINSQNGQNTITGLSNTLANVTNDKGSVTTQGN 1260  
DB 1201 rvgidqtkgltpkltvgnnngkgivinsngnqntitglSntlanvtnDkgsVttqgn 1260  
QY 1261 IKKEDKTRAAISIVDVLISAGNLOGNEAVDFVSTYDTYVNFANGNTTAKVYDDTSKTS 1320  
DB 1261 ikkedktraaisivdvvlisagfnlqngneavdfvstydytnfAngntTtakvYddtskts 1320  
QY 1321 KVVYDVNVDDTTIEVKDKKLGKTKTTLTSTGTGANKFALSNOATGDALVKASDIAVHLNT 1380  
DB 1321 kvvydvnvddttievkdkklgkvtktltstgtgankfalsnogatgdalvkaSdiavhlnt 1380  
QY 1381 LSGDITQAKGASOANNSAGYVDADGNKVIYDSTDNKYQAQNDGTVDKTKVAKDKLVAQ 1440  
DB 1381 lsgditqakgasoannsagYvdadgnkvIydstdnkyqaQndgtvDkTKvAkdklvaq 1440  
QY 1441 AQTDPGTLAQMNVKSVINKEQVNDANKKQGINEDNAFVKGLEKASDNKTKNNAVTVGDL 1500  
DB 1441 aqtpdgtlaqmNVkSVINkeQvNdankKqginEdnaFvkgLEkaSdnkTKnNAVtvgdl 1500  
QY 1501 NAVAQPTPLTFAGDTGTAKKLGCTLTIKGGQDTTNKLTDDNNIGVVAGTDGFTVKLAKDLT 1560  
DB 1501 navaqptpltfagdtgtakklgctltIk9gqdttnKltdnnigvVagtdgftvklakdlT 1560  
QY 1561 NUNSVNAGGTKIDEXGISFVDANGQAKANTPVLSEANGLDLGKVLISNVKGKTFKTDAAV 1620  
DB 1561 nUnsvnagGtkidexgisFvdangqakantpvlSangldlgkvlISnvkgKtfktdaaV 1620  
QY 1621 QQLNEVRNLLGLGNDNADGNQVNIADIKKDPNSGSSNRTVIKAGTVLGGKGNNDTEKLA 1680  
DB 1621 qqLnevrnllglgndnadgnqvnIadIkKdpnsGssnrtviKagtVlggkgnndteKla 1680  
QY 1681 TGGVQGVQDKDGNANGDLISNVVYKTKQDKGSKKALLATYNAAGQNTVYVNNPAAEDRINE 1740  
DB 1681 tggvqvGVQdkdgnangdlisnvvyKtkQdkgSkallatYnaagQntvYvnnpaaedrIne 1740  
QY 1741 QGTRFHFVNDGNGEPVVGQNGIDSSASCKSHVAICFOAKADGCEAAVAICGROTQAGNQSI 1800  
DB 1741 qgirfHFvndgngEPvvgqngidssasckshvaIcfOaKadgceaaVAicgrOtqagngsi 1800  
QY 1801 AIGDNAQATGDQSIATIGTGNVAVAGKHSAGTIDPSTVKADNSYSVGNNOFTDATQTDVFG 1860

Db 1801 aignhaqatgdsiaigtgnvvagkhsagldpctvkdnsysvgnnnqfdatcqtavfg 1860  
Qy 1861 VGNITVTESNSVALGNSAISAGTHAGTQAKKSDGTAGTTTATAGTGVKGFAGQTAVG 1920  
Db 1861 vgnitvtesnsvalgnsaisagthagtqakksdgttagttttagatgvtkgfagqtavg 1920  
Qy 1921 AVSVGASGAERRIQWAAAGEVSATSTDAVNGSQLYKATQSTANATNELDRIHONENKAN 1980  
Db 1921 avsvgasgaerriqwnaagevsatsstdavngsqlykatqstanaetnelidhrihonenkan 1980  
Qy 1981 AGISSAMAMASMPQAYIPGRSMVTTGGTATHNGOGAVAGLSKLSDNGOWKFKINGSADTQ 2040  
Db 1981 agissamamampqayipgrsmvtggtathngogavavaglsklsdngqvwfkingsadtq 2040  
Qy 2041 GHVGAAYGAGFHF 2053  
Db 2041 ghvgaavgagfhf 2053  
RESULT 2  
AAB69134  
ID AAB69134 standard; Protein; 2047 AA.  
XX  
AC AAB69134;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE M. catarrhalis strain 4223 genomic 200kDa protein SEQ ID NO:7.  
XX  
KW Moraxella catarrhalis strain 4223; major outer membrane protein;  
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection.  
XX  
OS Moraxella catarrhalis.  
XX  
XX W0200107619-Al.  
XX  
XX 01-FEB-2001.  
XX  
XX 26-JUL-2000; 2000W0-CA00870.  
XX  
XX 27-JUL-1999; 99US-0361619.  
XX  
XX (CONN-) CONNAUGHT LAB LTD.  
XX  
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;  
XX  
XX WPI: 2001-159722/16.  
XX  
XX N-PSDB; AAF59102, AAF59103.  
XX  
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
XX useful in protective vaccines and for diagnosis -  
XX  
XX Claim 1; Fig 3A-W; 247pp; English.  
XX  
XX The present invention describes an isolated and purified nucleic acid (I)  
XX that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.  
XX The 200 kDa outer membrane protein (II) has antibacterial activity and  
XX can be used in vaccines. (II), and its truncated versions, are used as  
XX immunogenic compositions and vaccines to protect against M. catarrhalis  
XX infections, particularly otitis media in humans. (II) is also used as  
XX antigen in immunoassays for detecting specific antibodies (Ab), and to  
XX generate Ab. (I) are used for recombinant production of (II) and its  
XX fragments are used as probes for identifying/cloning 200 kDa protein  
XX genes from other strains, and for diagnostic detection of M. catarrhalis.  
XX (I) makes possible production of large amount of recombinant immunogens.  
XX Expression of truncated versions of (II) reduces toxicity of the protein  
XX towards the Escherichia coli host. The present sequence represents the  
XX M. catarrhalis strain 4223 genomic 200kDa protein, which is given in the  
XX exemplification of the present invention.  
XX  
XX Sequence 2047 AA;

Query Match 74.1%; Score 7675; DB 22; Length 2047;  
Best Local Similarity 74.8%; Pred. No. 0;  
Matches 1604; Conservative 134; Mismatches 219; Indels 186; Gaps 28;  
Qy 1 MNHYKVFENKATGTFMAVEYAKSHSTGGSCATGQVGSVFTLSFARIAALAVLIGAT 60  
Db 1 mhykvvfnkatgtfmaaveyakshstggscatgqvgsctlsfariaalavligat 60  
Qy 61 LNSGAYAOQIITTK-TEIGOTNNKINNTLKGDLALATFASLKAOGSQAIAIGSKVP 119  
Db 61 lsgsayaqkdkkhaiaeqnqrrs--gtakaddraiaigenanadggqgaiaigssnk 118  
Qy 120 DPNNGSNGNVSNAKESIAIGGDVLAEGDASIAIGSDDLXL-----PKNLDLKN 170  
Db 119 tvngssldkigtatggesiaiggdvksasdasiaigsdldhldghgnpkpbg-tlin 177  
Qy 171 EFHKLHGHHEILKKTQSTGDKIKYRRTAQAQGHASTAVGAMSYAOGHFSNAGTATATA 230  
Db 178 d---linghavlkeirsskdnvdyrrttasghastavgamsyagqghfsnagftrataks 234  
Qy 231 AVSLAVGLAAOATKQSSIAVGSNAKANAFATAIGNTVNVNLGRGVALGFGSOILDRNN 290  
Db 235 ayslavglataegdstiaigsdatssslgailagatraglqgsiaigqsvvtqsdnn 294  
Qy 291 TDASAYVPLGKTADQYKATRGDSTDFISGNSNNSSIRRIINVGASRDTDAVNV 350  
Db 295 s-rpaytpntqalpkfcatntkagpl-sig-----snskrkiinvgagvntkdavnv 347  
Qy 351 AOLKLVLELA-NRKITFGDGNNSNVSVERGLNLTITKGDQTNALTEANTGVV--TDG 407  
Db 348 agleavvkwakeritfg--gdnstdvkigldntltikggaetnaltddnnlgvkeadn 405  
Qy 408 NGLVKVLAKELTGLTSV-----SATNKITVSTNNNAELQSGGLTFES-PITGTR-TDKT 460  
Db 406 sgikvlaktlnnltevnnttlnattvkvgssstaelldsltfcpntqsgstkl 465  
Qy 461 VVSDGLKFTNDSNSIATKGTTRITKKGIFAGTNGVDDESKPYLDNEKLKVGNSLTNSG 520  
Db 466 vygvngvkftnn----- 477  
Qy 521 SLTVNNTGNKQIQVGANGIKFATVANNVNTSAIVGTARITEEKIGFAGTNGVDDEQAP 580  
Db 478 -----aettaaigttrtrdkigfardgdvdekqap 508  
Qy 581 YLDKERLKVGRVEITDSSGINAGNHKITCLTNGIANTDAVTIKOLDKAKPTLNAGDGISI 640  
Db 509 yldkkqlkvgsvaltidngidagnkkslnlakgssandavtleqikaakptlnagagisv 568  
Qy 641 NSNNGDLVSSGNITPTTYNISVKTTKLNSNGTSGNNKFSVSNHNNLSVTAKLADLYL 700  
Db 569 tptelsvdaksgnvtcptynigvkteinsdgt--dkfsvkgsgtnnslvtaehlasyl 626  
Qy 701 NKVNETADSLPSFKVQNGDSEN-NAITVGKDT--NGKTFNTLKLKGENVNTNRTATG 757  
Db 627 nevnrtadsaigsftvkeedddanaaitvakdtknagavsilkgkngltvatkk-dg 685  
Qy 758 TVTFIDGDSNGLTTPKLVGSD-----TNGN-----RL 785  
Db 686 tvtfisgdsngltlqkstinnngdltvkdtnedqivgangikftnvnsgnpgtiantari 745  
Qy 786 VIEQV--PSADG-----NSTKNITKGLSPTLPSPASPSG 817  
Db 746 trdkigfagsgdvdkntkpyldqdklqvgnvkitntginagkaigtisptlpsiadqs 805  
Qy 818 RNIALGNITIEKDKSNASIDDLNAGFNMLKNGKDKDFVSTYDVFIDGNATATATVY 877  
Db 806 rnieigntiqdkdknaasindilntgfnlknnpidfvstydivdfangnattatvch 865  
Qy 878 DEANQTSKVAYDVNVDEKTIETLTGDKKKQLGVKTKLTETSTNGN-ATTF--STDDDDHA 934  
Db 866 dtanktskvvydvndvdtthitgtddnkkglgvkttklntksangnta nfnvnssdeda 925





Thu Sep 13 14:19:18 2001

```

XX PN W09634960-A1.
XX PD 07-NOV-1996.
XX XX 29-APR-1996; 96WO-CA00264.
XX PF 26-MAR-1996; 96US-0621944.
XX PR 01-MAY-1995; 95US-0431718.
XX PR 07-JUN-1995; 95US-0478370.
XX XX (CONN-) CONNAUGHT LAB LTD.
XX PI Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;
XX DR WPI; 1996-506162/50.
XX DR N-PSDB; AAT38740.
XX XX Moraxella outer membrane protein - useful as immunogen in protective
XX PT vaccine and for diagnosis
XX PS Claim 14; Fig 6; 109pp; English.
XX CC An approx. 200 kDa outer membrane protein (AAW04505) can be
XX CC isolated from Moraxella catarrhalis otitis media strain 4223
XX CC by electroelution, or expressed from a gene (see also AAT38740)
XX CC obtd. from a strain 4223 genomic library. Natural or recombinant
XX CC outer membrane protein is useful as an immunogen to protect
XX CC against infection by Moraxella, esp. M. catarrhalis. It can
XX CC also be used to detect antibodies, esp. for differential diagnosis
XX CC between bacteria that cause similar symptoms, and also useful as
XX CC a carrier for other antigens and used to raise antitumour
XX CC antibodies for conjugation to therapeutic agents.
XX SQ Sequence 1992 AA;

Query Match 71.5%; Score 7405; DB 17; Length 1992;
Best Local Similarity 74.2%; Pred. NO. 0;
Matches 1549; Conservative 135; Mismatches 218; Indels 186; Gaps 28;

Qy 56 VIGATLNGSAVAQITTK-TEIGQTNKINNTLKGDALATGEASTAFGSLSKAQSQAIAI 114
Db 1 migtatlsavayakddtkhiaieqgnprss--glakadgdratlaigenaagqgqai 58
Qy 115 GSVKPDPPNNGSNGVSHAKGNESIAIGDVIABGDASIAIGSDLLYL-----PKN 165
Db 59 gssnktvngssldkigtatgesialggdvkasgdasialgsadhlldqhgnpkhpq 118
Qy 166 LDLNKFEHLHILKIKIOTSDGKIKYRTRAOGHASTAVGAMSYAOGHFSNAGFTY 225
Db 119 -tclnd---linghavlkkeirsskndvkryrrttasghas cavgamsyagghfsna fgr 174
Qy 226 ATAAEAYSILAVGLAAQATVKQSSIAVGSNAKANAFAATAIGNTVYVNLGRVAGLFGSQIL 285
Db 175 akaksayslavglataaegqstiaigsdatsslgalaigagtraqlqgsialggsvvt 234
Qy 286 DRDNDTASAVYPLGKTLADQVKATQCGSDTDIFSIGNSNNNSIRKLIINVOGASRDT 345
Db 235 qsdnns-rpaytpntqaldpkfqatnntkagpl-sig-----snsikrkiinvagvntk 287
Qy 346 DAVNVAQLKVEELA-NRKITPFGDNNNSNVERGLNLTITIKGDAQTNALTEANIGV 404
Db 288 davnvaqlaavvkwakerritfq--gddnstdvkigldntitkgaetnaltddnngv 345
Qy 405 --TDGNGLVKLAKELTIGTSV-----SATNKITVNTNNNAELQSGGLTFS-PITGK 456
Db 346 keadnsglkkvkiaktlnnltevtnttlnattvkvgsssttaeillsdltftqptgsq 405
Qy 457 -TDKTVYSIDGLKFNDNSNIATKGTTRITKKKIGFAGTNDGVDSKPYLDNEKLVGNS 515
Db 406 staktyvgvngvkftnn----- 422

```

Db 1465 99qtdntkltannlgnvagtgtvkvklakdltnlnsvnagctklddkgvsfvdsggaka 1524  
 Qy 1589 NTPVLSANGLDGGKIVSNVSGTKDTPDAANVQQLNEVRLGLL---GNDNADGNQVNI 1645  
 Db 1525 ntpvlsangldggkivsnvgktdkdaavqqlnevrnlilglnagnadnagngvnia 1584  
 Qy 1646 DIKDPNSGSSNRVIVKAGTVLGGKGNNDTEKATGCGVQVDPKDGANGDLSNVVKT 1705  
 Db 1585 dikkdpnsgssnrtvirkagtvlggkgnndtektatggigvgvdkdngangdlnvvvkt 1644  
 Qy 1706 QKDGSKALLATYNAAGTNTVNNPABIDRINEOGIRFFHVNDGNOEPVVOGRNGIDS 1765  
 Db 1645 qkdgskallatynaagntvlnmpaealdrineeqgirffhvdngnqepvvvgngngids 1704  
 Qy 1766 SASGKHSVAIGFQAKADGEAAVAIGROTQAGNQSTAIQDNQATGDOSIAITGTVNVAGK 1825  
 Db 1705 sasgkhsvaigfqakadgeaavaigrqtqagngsaiqdnagatgddgsiaigtgnvvagk 1764  
 Qy 1826 HSGAIGDPSTVKADNSYSVGNHNFDTATQDFVCGVNNITVTSNSVALGNSAISAGT 1885  
 Db 1765 hsgaigdpstvkdnsysvgnhndftatqtdvfgvgnnitvtsnsvalgnsaisagt 1824  
 Qy 1886 HAGTOAKSKDCTAGCTTTTAGATGVKGFAGOTAVGAVSVGASGAEFRRTQNVAAGEVSATS 1945  
 Db 1825 hagtqakskdctagcttttagatgtvkgfagotavgavsvgsasgaerrlqnvaagevsats 1884  
 Qy 1946 TDVNGSOLYKATOSIANATNELDHRHQENKANAGISSAMAMSPQAYIPGRSMVTG 2005  
 Db 1885 tdvngsqlykatqsiatanatneldhrilhqenkanagissamamspqayipgrsmvtg 1944  
 Qy 2006 GIATHNGOGAVAGLSKLSNDGQWFKINGSADTQGHVGAAGAGFHF 2053  
 Db 1945 giathngogavagvlsklsndgqwfkingadtqghvgaavagagfhf 1992

## RESULT 5

AAB69137  
 ID AAB69137 standard; Protein; 1992 AA.  
 AC AAB69137;  
 XX  
 XX 24-APR-2001 (first entry)  
 XX  
 DE M. catarrhalis M56 200kDa protein in pK348 SEQ ID NO:13.  
 XX  
 KW Moraxella catarrhalis strain Q8; major outer membrane protein;  
 KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
 KW otitis media; detection.  
 XX  
 OS Moraxella catarrhalis.  
 XX  
 XX WO200107619-A1.  
 XX  
 XX 01-FEB-2001.  
 XX  
 XX 26-JUL-2000; 2000WO-CA00870.  
 XX  
 XX 27-JUL-1999; 99US-0361619.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Loosmore SM, Sasaki K, Yang Y, Klein MH;  
 XX  
 XX WPI; 2001-159722/16.  
 DR  
 DR N-PSDB; AAF59106.  
 XX  
 XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
 PT useful in protective vaccines and for diagnosis  
 PT  
 XX  
 PS Claim 1; Fig 8A-V; 247pp; English.  
 XX  
 XX The present invention describes an isolated and purified nucleic acid (I)

CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.  
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and  
 CC can be used in vaccines. (II), and its truncated versions, are used as  
 CC immunogenic compositions and vaccines to protect against M. catarrhalis  
 CC infections, particularly otitis media in humans. (II) is also used as  
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to  
 CC generate Ab. (I) are used for recombinant production of (II) and its  
 CC fragments are used as probes for identifying/cloning 200 kDa protein  
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
 CC (I) makes possible production of large amount of recombinant immunogens.  
 CC Expression of truncated versions of (II) reduces toxicity of the protein  
 CC towards the Escherichia coli host. The present sequence represents the  
 CC M. catarrhalis M56 200kDa protein in pK348, which is given in the  
 CC exemplification of the present invention.  
 XX  
 SQ Sequence 1992 AA;

Query Match 71.5%; Score 7405; DB 22; Length 1992;  
 Best Local Similarity 74.2%; Pred. No. 0;  
 Matches 1549; Conservative 135; Mismatches 218; Indels 186; Gaps 28;  
 Qy 56 VIGATLNGSAYAAQITTK-IEIGOTNKINNTLKGALATGEASIAFGSLSKAQSQAIAI 114  
 Db 1 migatlsysayaqkdkthiaigeqnprrs--gtakadgdraiaigenanaagggqatai 58  
 Qy 115 GSVKPDPPNNGSNGVSHAKNESIAIGDVLAEAGDASIAIGSDDLYL-----PRN 165  
 Db 59 gsnkntvgssldkigtatgesiaigdvkasdiaigsddhildqgnpkpkp 118  
 Qy 166 LDLNKFEHLHHEILKKIOTSTGKIYKRRTRAQGHASTAVGAMSYAQGHFSNAGFY 225  
 Db 119 -tlind---linghavlkerrsskdndvkrrttasghastavgamsyaqghfsnagfr 174  
 Qy 236 ATAEAAVSLAVGLAAQATKQSSIAVGSNAKAFATAIGNTVNLGRVALGFGSQTLL 285  
 Db 175 ataksayslaivglaaatkaegstiaigsdatsslgalaigagtragqlggsiaigggsvvt 234  
 Qy 286 DRDNNTASAYVPLGKTLADQYKATROGSDSTDFISGNSNNNNNSIRRKIIIVGAGSRDT 345  
 Db 235 qsdnns-rpaytpntqaldpkfqtatntkagpl-sig-----snsikrlinvagavntk 287  
 Qy 346 DAVNVAQLKVEELA-NRKITFKGDGDNNSNSVERGLONTLTIKGDAQTNAITEANIGV 404  
 Db 288 davnvaqleavvkwakerritfq--gddnstdkvlgldntlikggaealnaldnngvr 345  
 Qy 405 --TDGNGLVKVLAKELTGLTSV-----SATNKITVNTNNNAELQSGGLTFS-PTGTGK 456  
 Db 346 keadnsglkvklaklennlttevttnattkvkvgsssttaeillsdltftqtpntgsq 405  
 Qy 457 -TDKTVYSIDGLKFTNDSNSIATKGTTRITRKKIKIGFAGTNDGVDESKPYLDNEKLKVGNS 515  
 Db 406 stsktvgygvngvkftnn-----  
 Qy 516 TLNSGLTVNNTTGNKQIQVGANGIKFATVANNVANTSATVGTARITEEKTQAGTNDGV 575  
 Db 423 -----aettaagttritrdrkigfardgvd 448  
 Qy 576 DEQAPYLDKERLKYGRVEITTDGGINAGNHKITGTGTNGTANTDAVYIKOLDKAKPTLNAAG 635  
 Db 449 ekqapylckkqlkvgsvaitdngidagnkklsnlakssandavtieleqikaakptlnag 508  
 Qy 636 DGISINSNNGDLVDSSGNITPTTYNISVKTNTKLSNGTSGNNKFSVSNADHNSLVATKD 695  
 Db 509 agisvtpteisvdaksgnvtaptynigvkttelnsdglts--dkfsvkgsgttnslvtaeh 566  
 Qy 696 LADYLNKVNETADSAIPSKVQNGDNSN-NAITVGKDT--NOKTFTNTLKKGENGVNITTT 752  
 Db 567 lasylnevrntadsalqsftvkeedddanaaitvakdtknagavsiiklkgkngltvat 626  
 Qy 753 NRATGTVTFFGIDQSNGLTTPKLTVGSD-----TNGN----- 783  
 Db 627 kk-dgvtvfgisqsglgtigkstlnddglvtkdtnegiqvgangikfctvngnsnpptgla 685



```
QY 784 ---RLVTEQV--PSADG-----NSKNIKIGLSPTLPSI 812
Db      |: : : : :|
686 ntariIrdkIgfagsgdvndtkpyIdqdkIqvgnvkItntgInaggaItglspIpsi 745
QY 813 ASPSGRIALGNPIERKDSNAASIDDLVNLACGNLKNNGKDKDFVSTYDTYDFIDGNAT 872
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
746 adgsrnieIgtIqdkksnaasindIntgfnlknnpIdfvsyIdIvdfangnatt 805
QY 873 ATVTYDEANOTSKVAVDVNVDKTIETLTGDNKKQIQLGVTKITETSTNGN-ATTP--ST 929
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
806 atvthdtantkskvvydvndvddtIhtIgtddnkkIgvkttkInktsangntatnfnvs 865
QY 930 DDHAIIVKASDIAAGNLTLAEITHTTKGTANTALQTFVKKVVDENDKADDTNAITVKKDG 989
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
866 sdeIalvnakdiaenIntIakeIhtktgtadtalqIftvkkvdennadanaItvgqkn 925
QY 990 TSGKVNTLKLKGNGLDIDTKDGTVTFFGINTQSGLKAGDSTTLLNNGLSIKNTASNEQI 1049
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
926 anqvnItIikgngInIktkdgntvtfgInttsgIkagks-ItndgglIsiknptgseqI 984
QY 1050 QVGADGVKFMV--NNGVVGAGIDGTTTRIDETGFTGTSGLDKSPHLSKDCGINAGGKK 1108
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
985 qvgadgvkfakvnnngvvgagiDgtttrIdetgIftgtsGLDKSPHLSKDCGINAGGKK 1044
QY 1109 ITNIQSIEIAKNSHDVGTGKIYDLKTELENKISSAKTAQNSLHEFSVADEQGNFTVS 1168
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
1045 itniqsieIaqnshdvtgkIydlktelenkIsstaktagnslhefsvadeqgnftvs 1104
QY 1169 NPVSSYDTSKTSVDITPAGENGITTKVNGKVVVRVGDIDTFKGLTTPKLTGVNNGKGLIVN 1228
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
1105 npvssydtksdvtIfagengIttkvngkvrvvgIdgtkglIttkltvgvnnngkglvId 1164
QY 1229 SONGONTITGLSNTLANVTNDKGSVRTTEQNIKIDBDKTRAASIVDVLISAGFNLONGE 1288
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
1165 sqngntItlsntlanvtnksgsvrtteqnnIikdekttraasIvdlIsagfnlqng 1224
QY 1289 AVDFPVSYDVTNPFANGNTTAKVYVDFTSKTSKVYDVNVDDFTTIEVKDKKLGKVTTLT 1348
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
1225 avdfvstydtvnfagdnattakvtydydtsktskvvydvnnvddtIevkdkkIgvktlt 1284
QY 1349 STGTGANKFALSNGATDALVKASDIIVAHNLTLSGDITOTAKGASQANNSAGYVDADGNKV 1408
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
1285 stgtgankfalsngatgdalvKasdiIvahlntIsgdIqtakgasqannsgyvdadgnkv 1344
QY 1409 IYDSTDNKYYQAKNDGTVDTKEVAKDKLVQAQTPDGTTLAQMNVKSVINKEQVNDANKK 1468
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
1345 IydstdnkyyqakndgtvdktkEvakdkIvagaqtpdgttIaqmnnvksvinkeqvndankk 1404
QY 1469 QGINEDNAFVKLEKAAADNKTNAAVTVGDLNAVAOTPLTEFAGDTCTTAKKLGELTIK 1528
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
1405 qgInednafvKlekaasdnktnaavtvGdlnavagtpItfagdtgttakkIgetItIk 1464
QY 1529 GCQTDTNKLTDNNITGVVAGTDFGVTKLAKDLTLNLSVNAAGTKIDERKISGFVDGAOKA 1588
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
1465 gqqtDtnkltDnnIgvvagtDfvtkIakdltnlnsvnaagtkIdkgvsvfdssgqaka 1524
QY 1589 NTPVLSANGLDGGKVTNSVKGKTKDTPDAANVQQLNEVRNLLGL---GNDNADGNQVNI 1645
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
1525 ntpvlsangldlgvKsvngvkgtktdaannvgqInevrnlIlIagnadnagdnqvnIa 1584
QY 1646 DIKKDPNSGSSSRNTVTKAGTVLGGKGNNDTEKLTATGGVGVGVKDCGNANGDLNVVWKT 1705
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
1585 dIkKdpnsgsssrntvtkagtvlggkgnndLeklatgIqvgvdkdgnanagdlsvvwt 1644
QY 1706 QKDSKKALLATYNAAGOTNYVTNPPAIDRINEQIRFFHVNDGNQEPVQGRNGIDS 1765
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
1645 qKdgsKkallatynaagotnyltnpaeaidrineqgIrffhvnndgnqepvvggrngId 1704
QY 1766 SASGKHSVAIGFOAKADGEAAVATGRQTQAGNQSITATGDNAQATGDQSIATGTGNVAGK 1825
Db      |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|
1705 sasgkhsvaIgfoakadgeaaVaIgrqtqagngsIaIgdnaqatgdqdsIaIgtgnnvagk 1764
```

QY 1826 HSGAIGDFSTKADNSYSVGNNNQFTDATQTDVFGVGNNTVTESNSVALGSNSAISACT 1885  
Db |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|  
1765 hsgaIgdpsvkadnsysvgnnnqftdatqtdvfgvgnntvtesnsvalgsnsalsagt 1824  
QY 1886 HAGTQAKKSDGTAGTTTGTAGATGTVKGFAGQTAVCASVGSAGCAERRIQNVAAGEVSATS 1945  
Db |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|  
1825 hagTqakksdgtagttttagatgtvkgfagtcavgavsvgsagcaerrIqnvaagevsats 1884  
QY 1946 TDAVNGSOLYKATQSTANATNELDRIHONENKANAGISSAMAMASMPQAYIPGRSMVTG 2005  
Db |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|  
1885 tdaVngsqlykatqslanatneldhrihopenkanagissamamaempqayIpgRsmvtg 1944  
QY 2006 GIATHNGGAVAGVLSKJSDNQWVFKINGSADTQGHVCAAVGAGGFH 2053  
Db |: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|: : : : :|  
1945 giathngggavagvIskIsdnqgwVfkingsadtqghvgaavgagghf 1992

RESULT 6  
AAB69136  
ID AAB69136 standard; Protein; 2314 AA.  
XX  
AC AAB69136;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE M. catarrhalis les1 200kDa protein SEQ ID NO:11.  
XX  
KW Moraxella catarrhalis strain Q8; major outer membrane protein;  
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection.  
XX  
OS Moraxella catarrhalis.  
XX  
PN WO200107619-A1.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-CA00870.  
XX  
PR 27-JUL-1999; 99US-0361619.  
XX  
PA (CONN-) CONNAUGHT LAB LTD.  
XX  
PI Loosmore SM, Sasaki K, Yang Y, Klein MH;  
XX  
DR WPI; 2001-159722/16.  
XX  
DR N-PSDB; AAF59105.  
XX  
PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
PS useful in protective vaccines and for diagnosis  
XX  
PS Claim 1; Fig 5A-Y; 247pp; English.  
XX  
CC The present invention describes an isolated and purified nucleic acid (I)  
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.  
CC The 200 kDa outer membrane protein (II) has antibacterial activity and  
CC can be used in vaccines. (II), and its truncated versions, are used as  
CC immunogenic compositions and vaccines to protect against M. catarrhalis  
CC infections, particularly otitis media in humans. (II) is also used as  
CC antigen in immunoassays for detecting specific antibodies (AB), and to  
CC generate Ab. (I) are used for recombinant production of (II) and its  
CC fragments are used as probes for identifying/cloning 200 kDa protein  
CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
CC (I) makes possible production of large amount of recombinant immunogens.  
CC Expression of truncated versions of (II) reduces toxicity of the protein  
CC towards the Escherichia coli host. The present sequence represents the  
CC M. catarrhalis les1 200kDa protein, which is given in the exemplification  
XX of the present invention.  
XX  
SQ Sequence 2314 AA;

Query Match 69.7%; Score 7219.5; DB 22; Length 2314;

[illegible]

QY 1790 GROTQAGNSIAIGDQNAQTGDSIAIGTGNVNVVAKHSGAIGDPSTYKADNSYSGVNNQ 1849  
 Db 2051 grqtqagnsiaigdnagatgdsiaigTgnvvtgkhsaigdpstvkadnsysvgvnnq 2110  
 QY 1850 FTDATQDTPFGVGNNTVTESNSVALGSNSAISAGTHAGTQAKKSDGTAGTGTAGATCT 1909  
 Db 2111 fidatqdvfgvgnntvtresnsvalgsnsaisagthagctqakksdgtagtttagatgt 2170  
 QY 1910 VKFAGOTAVGAVSGASGAERRIQNVAAGEVSATSDAVNGSQLYKATQSTANATNELD 1969  
 Db 2171 vkfagotavgavsgasgaerriqnvaagevsatsstdavngsqlykatqggianatneld 2230  
 QY 1970 HRIHQENKANAGISSAMAMASHPQAYIPGRSMVTGGIATHNGOGAVAGLSKLSDNGQW 2029  
 Db 2231 hrihqenkanagissamamasmpqayipgrsmvtggtiathngogavavaglsksldngqw 2290  
 QY 2030 VFKINGSADTQGHVGAAGAPHF 2053  
 Db 2291 vfkingsadtqghvgaagagfhf 2314  
 RESULT 7  
 AAR99393  
 ID AAR99393 standard; Protein: 2353 AA.  
 AC AAR99393;  
 DT 15-JAN-1997 (first entry)  
 DE Haemophilus adhesion protein HA2.  
 KW Haemophilus adhesion protein.  
 OS Haemophilus influenzae type b strain C54.  
 PN WO9630519-AL.  
 PD 03-OCT-1996.  
 PF 22-MAR-1996; 96WO-US04031.  
 PR 24-MAR-1995; 95US-0409995.  
 PA (UYSL-) UNIV ST LOUIS.  
 PA (UNIW ) UNIV WASHINGTON.  
 PI Barenkamp SJ, St Geme JW;  
 PI WPI; 1996-455364/45.  
 DR N-PSDB; AAT41476.  
 XX Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in  
 PT vaccines against H. influenzae infection.  
 PS Claim 5; Page 66-73; 120pp; English.  
 CC Haemophilus adhesion protein HA2 (AAR99393) is associated with the  
 CC formation of surface fibrils involved in adhesion to various host  
 CC cells; it is also referred to hsf (Haemophilus surface fibrils).  
 CC Its amino acid sequence was deduced from a genomic DNA clone  
 CC (AAT41476) derived from Haemophilus influenzae type b strain C65.  
 CC Large quantities of recombinant HA2 can be produced in transformed  
 CC prokaryotic or eukaryotic host cells, for use in vaccines against  
 CC H. influenzae infection.  
 XX Sequence 2353 AA;  
 SQ

Query Match 11.6%; Score 1206; DB 17; Length 2353;  
 Best Local Similarity 23.8%; Pred. No. 4, 2e-45;  
 Matches 627; Conservative 295; Mismatches 853; Indels 856; Gaps 114;

QY 1 MNHYKVFIFNKATCTPMVAEYAKSHSTGGSCATQOVGSVRTLSFARIAALA----- 53  
 Db 1 mnklnfvnwmtgtwvvselttrhtk--rasatvetavlatillfatvganatedeel 58  
 QY 54 -----VLVIGA----- 59  
 Db 59 dpvvtapvlsfhsdkegtgekevtensnwgiyfdnkvikagaitlkagdnlikikntd 118  
 QY 60 -TLNGSAYAQQI-----TTKIEIG-----OTNKI 82  
 Db 119 estnassfysllkldltsavateklsfgangkvditsdanglklaktgngvnhngl 178  
 QY 83 NNTLKGDALATGEASIAFAGLSRAQGSQAIAIGSVKPDNPNNGNNGVSHAKN-ESIAI 141  
 Db 179 dstlpadvntngvlsstsftpndvekra--atvkdvlnagnlnkagktaggnvesvdl 235  
 QY 142 GGDVLAEGDASIAIGSDDL-----YLPKNLDLKNFHLIHIHHEILKK 184  
 Db 236 ---vsaynnvefigdkntldvitakengkttevktpktsvikedgklftgke--- 288  
 QY 185 IQTSTDGKIKYRTRRAOGHASTAVGAMSIAOGHFSNAGFYATAEAYSLAVGLAAQATK 244  
 Db 289 -nndtnkvt-----sntatdntdegng-----lvtakavld-avnkagwrvk 328  
 QY 245 QSSIAVGSNAKANAPATAIGNTVWNLGRVALGFGSQILDRDNTDA-----SAYVPL 299  
 Db 329 ttt-----angqngdfatvasgtnvtfesgdgtas-----vtdungngitvkydakvgd 379  
 QY 300 G-----KTLADQYKATROGDSTDIFSIGNNNNNNSIRKIIINVAGSGRDTDAVNAQ 352  
 Db 380 gikfssdkkivadtaltvtgk--vaeiakedd-----kklvnag----- 419  
 QY 353 LKLVEELANRKITPKGDGDN-----SNSVERGLGNTLTIKGDAQTNALTEANIGVVTD 406  
 Db 420 -dlvtalgnlswkakaeadtdgalegiskddevkagetvtfkag-----kulkvkd 470  
 QY 407 GNGLVKVLAKELTGLTYSATNKITVSNTNNNAELQSGLTFSPIITKTDKTVYSIDG 466  
 Db 471 ganfyslqdalgtlts-----itlggtt-----ggndaktvinkdg 508  
 QY 467 LKFTNDSNSIATGKTRITRKKIKGFAGTNDGDESKPYL---DNEKLVGKST---LN-- 518  
 Db 509 ltitpagnng-gttgtntisvtkdgikagnkaltnvvasglrayddanfdvlnmsatclnrh 567  
 QY 519 -----SGSLTVNNTGKIOIQVGANGIKPATVANNVANTSATVGTARITEEKIGF-AGTN 572  
 Db 568 vedaykglnlnehkanakplv-----tdstaatvgdlr-----klgwwvstk 610  
 QY 573 DGVEQAPYLDK--ERLKVGRVEITTDGGINAGNIHKITGLTNGIANF-----DAVTI 622  
 Db 611 ngtkesngvkqadevlftgagaatvtsksengkhti---tsvsaetkadcglekdgdcti 667  
 QY 623 KQLKDAKPTLNAGDGISINSSNGDLVDSSGNTITPTYNISVKTTLKLSNNGTSGNNKFSYS 682  
 Db 668 klkvdhqnt-----dnvltvgngngtavgkfet-----vktgatdad-----rgkvtk 712  
 QY 683 NAHDNNS---LVTAKDLADYLNKVNNETADSPFVKQGDNS---NNAITVGKOTNGKT 736  
 Db 713 datandadkkvatkvdvatain-----saatfvktenlttsaidednptdngkdalka 765  
 QY 737 FNTLKLK-GEN-----GVNIT-----NNRATGTVTFICIDQSNGLT---TPKLTVG 777  
 Db 766 gdtltfkaglnkvkrdgknitfdlaknlevktakvsdltitggnptggttatpkvnt 825  
 QY 778 SDTNGNRLVIEQVPSADGNSKTKNI-INGLSPTL--PSIASPSGRNLTALGNITIEKDKSNA 834  
 Db 826 stadglnfake---tadasgsknvylkgiatltitepsagakssh---vdlinvdatkksna 879  
 QY 835 ASIDDVLNAGFNLNKNGKDKDFYSTVDTVDIDGNATATTATVYDEANOTSKVAYDYNVDE 894  
 Db 880 asledvrlagwnldgngnnvdyvatydtvnftddstgtttvtvtq--kadgkgedwkiga 937  
 QY 895 KTIELTGDNCKKQLGVKTKIKLTETSTNGNATFTFSTDDH-----ALVKASDI----- 941

Db 938 ktsvikhngklftg-kolk-----danggatvseedgdkdgtgltvaktavidavnsksw 991  
Qy 942 -----AGNLNTLAE--EIHTRTKTANTALQTFVTKKVDEN----- 974  
Db 992 rvtgegataetgatavnagatvtsvfnkgntatta-----tvsdkdngninvkydvn 1047  
Qy 975 -----DK--ADDTNATVNGDKTSGKV-----NTLKLKKNGL----- 1005  
Db 1048 vgdglikgdgdkkivadttltv-----tgkvsvpagansvnnkklvnaeglatalnlns 1103  
Qy 1006 -DIKTDKGTFTFGINTOSGLKAGDSTTLN--NGLSIKN-----TASNEQIQVGADGVKE 1058  
Db 1104 wtakadkyadesegedqevkagdkvfkagknklykqskdfyslqdtlgtltsltl 1163  
Qy 1059 AMVNVGVGAGIDGTRITRDEIGFTGTNGS---LDKSKPH---LSKDCINAGGKKTINI 1112  
Db 1164 ggtangrndtg---tvinkdgltitlangaaagtdaangntisvtdkagisagnkeilnv 1219  
Qy 1113 QSG-----EIAKNSHDAVGGKTYDLKTELENKISSTAKTAQNSL 1152  
Db 1220 ksalikytdntadqtdqkfehaavkanevefvgk-----ngatvsaktcdngk 1270  
Qy 1153 H-----EFSVADEQGNFTVSNPYSSYDTSK-----TSDVI 1183  
Db 1271 hvtldvaeakvgdglekddgkiklkdvndtgnnlltvdatkasvakgefnavttat 1330  
Qy 1184 TFAENG-----ITTKVNGKVVVRGIDQTKGLT----- 1211  
Db 1331 taqgtnanerkvvgkngngatataetdkkvatvgdvakaindaatfvkvenddeatidd 1390  
Qy 1212 -----TPKLTGVGNNG 1222  
Db 1391 sptddgandalagdtltlkagknkvrkgdnitfalandsvksatvsdklsigtgn 1450  
Qy 1223 KGIVINSQNGN-----TINGLSNLTANVTNDKGSVPTTEQGNLIIRKDETR 1269  
Db 1451 kvnltsdtkglnfakdsktgdanlhngiaslttdtllnsgat-tlgnngitdnekr 1509  
Qy 1270 AASIVDLVSAGFNLOG-----NGEAVDFVSTVDTVFANGNTTTAKVTYD--DTSKT 1319  
Db 1510 aasvkdvlmagvrvkvpasannqvenidfatvdytdvfygdkdtsvtveskdngkr 1569  
Qy 1320 SKVYDVNVDDTTIEVKDKKLGVKTTTTSFG-----TGAN-KFALSNOAT-----G 1365  
Db 1570 tev-----kigaktsvikhngklftgkelkdannngvtvtetdgkdeg 1613  
Qy 1366 DALVKASDIVAHLNPLSGDIOTAKGASOANSAGYVD-----ADGNKVIYDSTDNKYY 1418  
Db 1614 nglvtakavidavnkagvrvtkt-gangqnddfatvasgtnvtfadngttaevt----- 1667  
Qy 1419 QAKNDG--TVDKTRVAK-----DKLVAQA---QTPDGTILAQMNVKSVINKEQVNDANK 1467  
Db 1668 -kandgsitvkynkvadgklldgkivadttvltvaqdkvtapn-----ngdgk 1716  
Qy 1468 KOGINEDNAFVKGLEKAASDNKTKNAAVTVGDLNNAVOTPLTFAGDGTAKKLGTELTI 1527  
Db 1717 k-----fvdasgladalnklswa-tagkegtgevdpsnsgq-----evkagdkvtf 1763  
Qy 1528 KGGOTDTNKLTDNNIGVVAGTDGFTVKLAKDLTNLSV-----NAG-----GTTKIDEGIS 1578  
Db 1764 kag-----dnlikqsgkdfyslkkelkdltsvefkdanggtgsestklckdgl 1814  
Qy 1579 FVDANGQA-----KANTPVLASGLDLGGKVTISNVGKGT-----DTPDAAN 1619  
Db 1815 itpangagaagantantitsvtdkdisagnkavtnvsvglkfgdghtlangtvadfehy 1874  
Qy 1620 VOQLNEVRNLLGLGNDN-----ADGNQVNIAD-----IKKDPNSGS-----SSN 1658  
Db 1875 dnaykdllnldekagdnptvadntaatvgdlrglwwisadkttgepnqeynaqvnrnan 1934  
Qy 1659 RTVTKAGT-----VLGG-----KGNNDTEKLATGG-----VQVG----- 1687

Db 1935 evkfksngingvsgktlngtrvitfelakgevvksneftvknadsetnlvkvgdmyysk 1994  
Qy 1688 VDKDGNAGDLSNVVWVKTK-----DGSKALLIATYNAAGQTNVYVNNPAAEID 1736  
Db 1995 edidpatskpmtg---ktekykvengkvvsangsktevtltknkgsq---yvtnq--qvad 2046  
Qy 1737 RINEGIRIFFHVNDGNQEPVVOGRNGIDSSASGKHSVAIGFOAK-----ADGEA 1785  
Db 2047 alaksgef-----gladaeaeakfaesakdkqlskdkaetvnaahkv 2090  
Qy 1786 AVAIGROTQAGNQSTIAIGDNAOATGDO-----SIAIGTCNVVAGKHSIGAIGPSTVK 1837  
Db 2091 rfanglnlckv---saatvestdangdkvttfvtkdvelplqlty---ntdangnkivkk 2144  
Qy 1838 ADNSYSVGNNOFTDATQTDVFGVGNNTITTESNSVALGSNSAISAGTHAGTQAOKSDGT 1897  
Db 2145 adgkv-----yelnadgtasnkevtlgnvdangkk-----vvkvteq 2182  
Qy 1898 AGTTTATAGATVKGFGAGTAVGAVSVG-----ASGAERRIQNVAAGEVSATS 1945  
Db 2183 adkwyytnadgaadktkgevsndkvstdekhhvridpnngsqngkvvidnvangeisats 2242  
Qy 1946 TDAVNGSOLY---KATQSTANATNELDHRHONENKANAGISSAMAMASMPQAYIPGRSM 2002  
Db 2243 tdaingqliyavakgvtlnlagvnnlegkvkvkradagtasalaasqlpqtatmpgksm 2302  
Qy 2003 VTGGIATHNGQCAVAVGLSKLSDNGQWVFKINGSADTQGHVGAAGVAGPHF 2053  
Db 2303 valagssyvgqnglaigvrisdngkviirlsgttnsqgktvaagvgvq 2353  
RESULT 8  
AAB23860  
ID AAB23860 standard; Protein; 2411 AA.  
XX AAB23860;  
AC AAB23860;  
DT 17-JAN-2001 (first entry)  
XX  
DE Haemophilus influenzae adhesin (Hia) protein from type c strain API.  
KW Hia; adhesin; Haemophilus influenzae adhesin; NTHI; infection; vaccine;  
KW non-typeable Haemophilus influenzae; antinflammatory; auditory;  
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;  
KW diagnosis; immunogenic; antigen.  
OS Haemophilus influenzae.  
XX  
PW WO2000055191-A2.  
XX  
PD 21-SEP-2000.  
XX  
PF 16-MAR-2000; 2000WO-CA00289.  
XX  
PR 16-MAR-1999; 99US-0268347.  
XX  
PA (CONN-) CONNAUGHT LAB LTD.  
XX  
PI Loosmore SM, Yang Y, Klein MH;  
XX  
DR WPI; 2000-618897/59.  
DR N-PSDB; AAA92499.  
XX  
PT Novel nucleic acid encoding Haemophilus influenzae adhesin protein, for  
PT use as antigens and vaccines and for treating Haemophilus influenzae  
PT infection  
XX  
PS Claim 1; Fig 24; 275pp; English.  
CC The present sequence represents a Haemophilus influenzae adhesin  
CC (Hia) protein from the type c Haemophilus influenzae strain API.  
CC Hia genes and proteins have antinflammatory, auditory and antibacterial  
CC activities, and can be used in the production of a vaccine. An

CC immunogenic composition comprising an Hia gene, a polypeptide encoded  
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
CC protection against disease caused by Haemophilus strains in a  
CC susceptible host, preferably a human. An Hia protein is useful as an  
CC antigen, in immunogenic preparations including vaccines, as a carrier  
CC for other immunogens, and in the generation of diagnostic reagents. Hia  
CC is useful for treating diseases caused by the infection of Haemophilus  
CC influenzae such as meningitis, epiglottitis, septicaemia and otitis  
CC media. Recombinant production of Hia favours high recovery of the  
CC protein compared to the low recovery of native protein from Haemophilus  
CC influenzae species. A truncated protein has a significantly higher  
CC amount of recovery than a full-length protein.

XX Sequence 2411 AA;

Query Match 11.6%; Score 1205; DB 21; Length 2411;  
Best Local Similarity 23.8%; Pred. No. 4.8e-45;  
Matches 630; Conservative 299; Mismatches 863; Indels 880; Gaps 115;

Qy 1 MNHIYKVFNFKATGTFMAVAEYAKSHSTGGSCATGQGVSVRTLSPTAALAA-----53  
Db 1 mnkifnvwmtqtvvvseltrthck--rasatvetavlatllfatvganadedeel 59  
Qy 54 -----VLVIGA-----TLN 62  
Db 59 dpvvrtpvlsfhadkegtgekevtenwnwgiyfhngkvglkagaitlkagdnlikqstn 118  
Qy 63 GSAAQAQI-----TPIKIEIG-----QTNKNINNTL 86  
Db 119 assfysikkldltsvateklsfsgangkvditsdanglklaktngnvhngldstl 178  
Qy 87 KGDALATGEASTFGSLKAGQSOAIAIGSVKPPNNGSNGVSGSHAKGN-ESTIAIGDV 145  
Db 179 pdavntvgvslssstfpndvektra--atvkdlngwnlikgaktagnvsvdvl---v 232  
Qy 146 LAEGDASIAIGSDDL-----YLPKNLDLANEPFKLIHGHIEILKKIQTS 188  
Db 233 saynnvefitgkntldvltakengkttvfktpktsvikekdgklftgke-----nn 286  
Qy 189 TDGKTKYRTRAGHASTAVGAMSAQGHFSNAFGTYATAEAYSLAVGLAAQATKOSSI 248  
Db 287 dlnkvt-----sntatdnatdegng-----lvakavid-avnkagrvkttt- 327  
Qy 249 AVCSNAKANAFATAIGGTVVNLGRVALGFGSOILDRNNTDA-----SAYVPLG--- 300  
Db 328 ---angngdfatvasgntvtfesgdgtas-----vktngngitvkdyakvgdglkf 379  
Qy 301 ----KTLADQYKATROGSDTDIFSGNSNNNSIRRKTIINVGAGSRDTDAVNVAAQLKLV 356  
Db 380 dsdkkivadtaltvtgk--vaeiakedd-----kkklvnag-----dlv 418  
Qy 357 EELANRKITFKGGDNNNSNVERGLGNTLTIKGAQTNALTEANIGVYVDGNGLKVLIK 416  
Db 419 talgnlswkakaeadtdtdgaleglskdqevkagetvtfkagknkvkdqganfyslqd 478  
Qy 417 ELTGTSVSATNKITVSNTNNNAELQSGGLTFSPITGKTOKTVYSIDGLKFTNDSNSI 476  
Db 479 altglts-----itlgtttn-----gngdaktvinkdglittpagng- 515  
Qy 477 ATKGTTRITKKIKGFAGTNDGVDESKPYL---DNEKLKVGNSP---LN-----SGSLT 523  
Db 516 gttgtntsvtkdkigakgnkainvasglsrayddanfdvlnnsatdlnrhvedaykglln 575  
Qy 524 VNNTTGNKIQVQANGIKFATVANNVANTSATVGTARTITEEIKGF-ACNTNDGVDEQAPYL 582  
Db 576 lneknanqplv-----tdstaetvgdlr---klgwvstknngkteesngv 618  
Qy 583 DK--ERLKVGRVEITTDGGINAGNHKTTGLTNGLIANT-----DAVITKQLKDAKPTL 632  
Db 619 kqadeviftgagaatvtsksengkhti---tvsvaetkadsglegdgtklkvdnqnt- 674  
Qy 633 NAGDGISINSNNGDLVDSNGITTPYTNISVTKTKLNSNGTSGNNKFSVSNADHNNS--- 689

Db 675 ---dnvltvgngtavtkgget-----vktgatdad-----rgkvtvkdandadkk 720  
Qy 690 LVYAKDLADLYLNKVNKNETADSALESPFKVQNGDNS---NNALTVCKDNTNGTFTNTKLK-GE 745  
Db 721 vatvkdvatain-----saatfvktenltsidednptdngkddalkagdtltfkagk 773  
Qy 746 N-----GVNIT-----TNRATGTVTFIGDOSNGLT---TPKLTVGSDTNGNRLVI 787  
Db 774 nlkvkrdgnkntfclaknlevktakvsdttlignptggttatcpkvnitstadgnfak 833  
Qy 788 EQVPSADGNSTKNT-INKLSPTL--PSIASPSGRNALGNTIEBKDKSNASISDDVNLNAG 844  
Db 834 e---tadasgsknvykgiattliltepsagakssh---vdlinvdatkksnaasiedvlrag 887  
Qy 845 FNLXNCKGDKDFSVSYDTYDFIDGNATTATVYDEANQTSKVAYDVNVVDKTIETLFGDNG 904  
Db 888 wllgngnnvdyavdydlvnftddstgtttvtvtq--kadgkgadvgkigaktsvikhng 945  
Qy 905 KQOLGVKTIKLTETSTNGNATFTSTDDH-----ALVKSADI-----941  
Db 946 klftg-kalk-----danngatvseddgtgtgtglvtaktvldavnksgwrvtgegetae 999  
Qy 942 -----AGNLTIAE--EIHFTTKGTANTALQTFVKKVDEN-----D 975  
Db 1000 tgatavngnaetvsgtsvfkngnatta---tvskdngnlnvkydvnvgdglkigd 1055  
Qy 976 K--ADDTNATVVGKDGTSKGV-----NTLKLKCKNGL-----DIKTDKGT 1014  
Db 1056 kklvadttltlv-----tgkvsvpagansvnnkklvnaeaglatalnlnslwtakadkyad 1111  
Qy 1015 VTFGINTQSLKAGDSTTLN--NGLSIKN-----TASNEIQVQAGDGVKFAMVANGVGA 1068  
Db 1112 gesegtdgdevkagokvtfkagknkvkqsekdfyslqdtltgltstlgttgangrndt 1171  
Qy 1069 GIDGTRIRDRIGFTGTNGS---LDKSKPH---LSKGINAGKKIINIOS-----1114  
Db 1172 g---tvlnkgdltitlangaaagtdaangntisvtkdgisagnkeifnvksalktykdt 1227  
Qy 1115 -----GETAK-----NSHDA-----VTGGKIYDLK-----1134  
Db 1228 qntagatqpaantaevakqdlvdltkpatgaagngadadakpdttaatvgdglglgwisa 1287  
Qy 1135 -----TELE-----NKISSTAKTAQNSLH-----1153  
Db 1288 kktadetqkhefaavknaevefvgkngatvsaktddngnkhtvtldvaeakvgdglekd 1347  
Qy 1154 -----EFSVADEOGNNFTVSNPYSSYDTSK-----TSDVITPAGENG-----ITTKVNM 1197  
Db 1348 tdkkiklvndtgdgnllvtvdatkgsavakgefnavttatlaagtnanergkvvvkgsn 1407  
Qy 1198 GVVVRVGIDQTPKGLT-----1211  
Db 1408 gatateldkkkvatvgdvakalndaatfvkavenddaatidsptddgandalakagdtl 1467  
Qy 1212 -----TPKLTVCNNNGKGVINSONGN-----1234  
Db 1468 kagknkvkrdgnkntfalandsvksatvsdklsigcngknknitsdtkglnfakdskt 1527  
Qy 1235 -----TITGLSNTLANVNDKGSVRTTEQGNIIKDEKTRAASIVDVLISAGNLOG--- 1285  
Db 1528 gddanhlnglasltltdllnsgat-cnlgngitldnekkraasvkdvlvlnagvngvfkp 1586  
Qy 1286 -----NGAVDFSVTYDTVPNFANGNTTTAKVTD--DTSKTSKVYVDVNVDDTTIEVKDK 1338  
Db 1587 asannqvenlfdvatydvtfvsgdkdtvavveskdngkrtev-----1630  
Qy 1339 KLGKVTTLTSTG---TGAN-KEALSNAQT-----GDALVKASDIVAHLNLTUSD 1384  
Db 1631 klgatksvikdngklftgkelkdannngvtvtetdgdgdegnglvtakavidavnkagwr 1690  
Qy 1385 TOTAKGSAQNNAGYVD-----ADGNKVIYDSTDNKYQAOKNDG--TVYDKTKEVAK- 1434

Db 1691 vktk-gangqnddfatvasgtnvtfadngngttaevt-----kandgsitvkynvkvadg 1743  
 QY 1435 -----DKLVAQA---QTPDGTQAMNYSKVINKEQVNDANKKOGINEDNAPVKGLEKAAS 1486  
 Db 1744 lkldgdkivadttltvadvgtkvtapn-----ngdgk-----fvdasglada 1785  
 QY 1487 DNKTNAAVTVGDLNVAQPLTFAGDTGTAKKLGETLTIKGGQDTNKLTDNNIGVVA 1546  
 Db 1786 lnklswta-tagkegtgevdpsagsg-----evkagdkvtfkag-----dnlkikg 1831  
 QY 1547 GTDGTFTVKLAKDLTNLSV-----NAG-----GTRKIDKESFVDANGQA-----KANTPV 1592  
 Db 1832 sgkdfysllkkelkdlsvfkdanggtgsestkitkdgititpanagagaagantatis 1891  
 QY 1593 LSAAGLDGKGVISNKGKTK-----DPTAANVOQLNEVRNLLGLGNDN-- 1636  
 Db 1892 vtdgisagnkavtnvsglkkfgdghtlangtvadvfdekhynaykdltnldekagdnnp 1951  
 QY 1637 --ADGNQVNTAD-----IKKDPNSGS-----SSNRTVIKAGT-----VLGG 1670  
 Db 1952 tvadntaatvgdlrglgwvisadkttgepngeynagvrnanevkfsgnginvsgkting 2011  
 QY 1671 -----KGNNDTEKLTATG-----VOVG-----VDKDGANGDLSNVVVKTO 1706  
 Db 2012 trvitfelakgevvksneftvkadgsetnlkvvgdmyskediidpatskpmtg---kte 2068  
 QY 1707 K-----DGSKALLATYNAAGQTYNTNNPAAEADIRNEOGIRFFHVNDGNQEP 1755  
 Db 2069 kykvengkvsvsangsktevtltnkgsq---yvtgn--qvadaiaaksgfel----- 2113  
 QY 1756 WQGRNGIDISASGKHSVAIFGQAK-----ADGEAANAIGRTQNGNSIAIGD 1804  
 Db 2114 -----gladaaeakafaesakdkqiskkaetvnaahdkvrfanglntkv---saatve 2164  
 QY 1805 NAOATGQ-----SIAIGTGNVAGKHSAGTGPSTVKADNSYSVGNNNQFTDATQF 1856  
 Db 2165 stdangdkvttfvktadvelpitqiy---ntdangnkivkkadgkw-----yel 2210  
 QY 1857 DVFGVGNNTVTESSVALGSNSAISAGTHAGTQAKSDGTAGTGTAGTGTGKVPAGQ 1916  
 Db 2211 nadgtasnkvtlgnvdangk-----vvkvengadkwyytnadgaadkckge 2259  
 QY 1917 TAVGAVSVG-----ASCAERRIONVAAGEVSATSTDAVNGSOLY---KATQSI 1961  
 Db 2260 vsndkvstdekthvrrldpnnqsgkgvvidnvangeisatsatdaingsqlayavakgvtln 2319  
 QY 1962 ANATNELDHRTHQENKANAGISSAMAMAPQAYIPGRSMVTGGIATHNGOGAVAVGLS 2021  
 Db 2320 agqvnnlegkvkvgradagtasalaasqlpqatmpgksmvalagssyqqnglaigvs 2379  
 QY 2022 KLSDNQGVWFKINGSADTQGHVGAAGVAGPHF 2053  
 Db 2380 risdngkviirlsgttnsqgktgvaagvgvqw 2411

RESULT 9  
 AAB23856  
 ID AAB23856 standard; Protein; 1104 AA.  
 AC AAB23856;  
 XX  
 XX 17-JAN-2001 (first entry)  
 DT  
 XX  
 DE Haemophilus influenzae adhesin (Hia) protein from NTHi strain 29.  
 KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;  
 KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;  
 KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;  
 KW diagnosis; immunogenic; antigen.  
 OS  
 XX Haemophilus influenzae.  
 XX  
 PN W0200055191-A2.

XX 21-SEP-2000.  
 XX 16-MAR-2000; 2000WO-CA00289.  
 XX 16-MAR-1999; 99US-0268347.  
 XX (CONN-) CONNAUGHT LAB LTD.  
 XX Loosmore SM, Yang Y, Klein MH;  
 DR WPI: 2000-618897/59.  
 DR N-PSDB; AAA92495.  
 XX  
 PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for  
 PT use as antigens and vaccines and for treating Hemophilus influenzae  
 PT infection  
 XX  
 PS Claim 1; Fig 20; 275pp; English.  
 XX  
 CC The present sequence represents a Haemophilus influenzae adhesin (Hia)  
 CC protein from the non-typeable Haemophilus influenzae (NTHi) strain 29.  
 CC Hia genes and proteins have antiinflammatory, auditory and antibacterial  
 CC activities, and can be used in the production of a vaccine. An  
 CC immunogenic composition comprising an Hia gene, a polypeptide encoded  
 CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
 CC protection against disease caused by Haemophilus strains in a  
 CC susceptible host, preferably a human. An Hia protein is useful as an  
 CC antigen, in immunogenic preparations including vaccines, as a carrier  
 CC for other immunogens, and in the generation of diagnostic reagents. Hia  
 CC is useful for treating diseases caused by the infection of Haemophilus  
 CC influenzae such as meningitis, epiglottitis, septicaemia and otitis  
 CC media. Recombinant production of Hia favours high recovery of the  
 CC protein compared to the low recovery of native protein from Haemophilus  
 CC influenzae species. A truncated protein has a significantly higher  
 CC amount of recovery than a full-length protein.  
 XX  
 SQ Sequence 1104 AA;

Query Match 9.9%; Score 1021; DB 21; Length 1104;  
 Best Local Similarity 27.5%; Pred No. 2.3e-37;  
 Matches 363; Conservative 163; Mismatches 402; Indels 390; Gaps 50;

QY 872 TATYTYDEANQTSKVAYDVNVDEKTIELTGDKGKQLGVKT-----IKLTETSTNGNA 924  
 Db 141 talsataeanntsvtnglnaygdtfnfntmnsiadlekhvqdaykgilnlnekt--nk 98  
 QY 925 TTFSTDDHALVKASDIAGNLTAEIHTTKGTANTALQTFTVKKVDENDKADDTNAT 984  
 Db 99 ssflivadntaat-----vgnlrklgwvlsskngtrne--ksygvkqade----- 140  
 QY 985 VGKDGTSCKVNTLKLKGNGLDI-KTDKDGTVTFGINTQSLKAGDSTLNNGLSIKNT 1043  
 Db 141 -----vifdsgaatvssskdgkhtltisvksfaevkcdattg----- 182  
 QY 1044 ASNEQIOVGADGVKFAVMNNGVWGAGIDGTTTRITRDEIGFTGNGSLDKSPHLSKDGIN 1103  
 Db 183 -----gvnadrgvkaeden--gadvd----- 202  
 QY 1104 AGGKRTNIQSGEITAKNSHDVATGCKIYDLKTELENKTSSTAKTAQNSLHFEFSVADEQGN 1163  
 Db 203 ---kkvatvk--dvakaindaatfvkvestdddiengaagknettdqal----- 246  
 QY 1164 NFTVSNPSSYDTSKTSPTVITFAGENGITTKVNGK-----VVRVGIDQTKGLTTPKLT 1217  
 Db 247 -----kagdtlikagknlkakldqngkvsftalaxkldvtsakvsdklsi 292  
 QY 1218 GNNNGK-----GIVI-----NSONGQNTITGLSLTANVTNDKGSVTTTEOGNIK 1263  
 Db 293 gkdtknkvitsdangklaktngngqngvnhngiaastl----tditgnttqasgva 348  
 QY 1264 DEDTRAASIVDVLISAGFNQNGEAVDFVSTYDTPNFANGNTTTAKVTYDDTSKTSKV 1323

Qy	872	TATVTVDEANTGSKVAYDVNVDEKTIETDGGKKQLGVKT-----IKLTETSTNGNA	924
Dp	41	talsataeanntstnglnaygdncfnatnnsiadlekhhvqdaykglmlnekd--nk	98
Qy	925	TTFTDDDHALVKASDIAGNLMTLAEETHTKGTANTALQTTFVKKVDENDKADDTNAIT	984
Dp	99	ssflvadntaat-----vgnlrklgwilsskngrtne---ksyvqvkade-----	140
Qy	985	VKGDTSGKVNTLKLGRNGLDI -XTDKDGTTFVGINTQSGLKAGDSTTLLNNGLSIKNT	1043
Dp	141	-----vlftgsgaatcvsssskogkhltitlsvtkgsfaevktdattgg-----	182
Qy	1044	ASNEQIOVGADGKVFAMVNWVGVGACIDGCTWTRTRDRDEICGTFGTNGSLDKSPHLSKDGIN	1103
Dp	183	-----qvnadrgkvkaeden--gadvd-----	202
Qy	1104	AGCKKITNIOSGEIAKNSHDAVTGGKIYDLKTELENIKISSTAKTAONSILHEFSVADEQGN	1163
Dp	203	---kkvatvk--dvakalindaatfvkvestdddiengaagknettdqa!-----	246
Qy	1164	NFTVSNPYSYDFPSKSDVITFAGENIITKVNKG-----VVRVCIDOTKGLTTPKLTAV	1217
Dp	247	-----kagdtitlkagknklkaidqngksvtfalakdidvsckvsokls!	292

QY	1044	ASNEQIQVGADGVKFAWVNVGVGAGIDGTTTRTRDEIGCTGTNGSLDKSKPHLSKDGIN	1103
		:      :	
Db	183	-----qvnadrgkvkaeden--gadvd-----	202
QY	1104	AGGKKTINTQSGEIAKNSHDVAVTGGKLYDLKTELENNKISSTAKTAONSLHFESVADEQGN	1163
		: :    :      :    :   :	
Db	203	---kkvatvk--dvakalndaatfvkvestdddiengaagknctdqaI-----	246
QY	1164	NETVSNPYSSTKTSKTSVDVITFAGENGITTKVNGK-----VVRVGIDGTGKLTTPKLTAV	1217
		:      :      :      :	
Db	247	-----kagdtlIkagknIkakIQdgkvsytfalakldvtskavksokIsI	292





QY 1325 DYNVDDTTIEVKKLGKVKVTTTTSTGTGANKFALSNOATGDALVKASDIVAHLNTLSGD 1384  
Db 302 -rvdvgtipvq----- 312  
QY 1385 IOTAKGASQANNSAGYVDADGNKYIYDSDNKKYQAKNDGTIVDKTEYAKDKL----- 1437  
Db 313 -----yvtdgetv--kvgnyyeakdgsadmddkvgengklaktkvkl 355  
QY 1438 -----VAQQTDPGTLAQNVKSVINKQVNDANKKGINEDNAFV-----KG 1480  
Db 356 vsangtnpvkslnvadgt---entdavsfkqlkalqdkvltlsasnayangsdadgkg 412  
QY 1481 LE-----KAASDNKT-----KNAAVTVGD----- 1499  
Db 413 iqltlnlgnfkfstgdgellnikaedntvftpkpksvqvgdgdgkatidgaktgttgive 472  
QY 1500 -----LNAVQTPPTTFAGDTGTTAKKLGELTTLTKGGQTDNKLTDNNIGVVGAGTDGFT 1552  
Db 473 aselvdslnklgwkvgtgdgtgvtgdthdtl-vksgdvltlkagdnlkvkqegtn-ft 530  
QY 1553 VKLAKDLTNE-NSV-----NAGGTKIDKIGISFVDANGQA-----KANTPVLISANCL 1598  
Db 531 yalldeltdvksvefkdtangangastktkdglittpangagaagantantistvktkgd 590  
QY 1599 DLGGKVISNVGKTKTDANVQQLNEVRNLGLGNDNADGNQVNIADIKDPNSGSSN 1658  
Db 591 sagnkavnvsvglkkfgdanfopltssadnltkydyndaykgtlnldexskgkqtpvad 650  
QY 1659 RTVTKAGTVLGGKGNNDTEKLTATGGVGVGDKDGNGANDLSNV-WYKTKDGSKKALLAT 1717  
Db 651 nt-----aatv-----gdlrglgwv-isadkktkgeinke 678  
QY 1718 YNAAGQTNVYVNNPAEALDRINEQGIREFPHVNDQNPVQGRNGIDSSASGK-----HSV 1773  
Db 679 yna-----qvrnanevfkfsg-ngl--nvsgkltldngtr 709  
QY 1774 AIGFOAKADGEAAVAIGROTQAGNQSIAIGDQNAATGDSIAIGTGNVAGKUSIGAIGDP' 1833  
Db 710 eitfelakdena-----lafgsgskalrdntvaigtgnvnaeksgafgdp 755  
QY 1834 STV---KANNSVSGVNNQFTDATQDVFVGVCNNI-----TVTESNSV----- 1873  
Db 756 nyiedkaggsyafgndnri---tskntfivgnsvnaekrdangnvitkeevvgkdgaktk 812  
QY 1874 -----ALG---SNSAISAGTHAGTQAK-----KSDGTAGTTTACATGTVKGFAGQTAVGA 1921  
Db 813 vlvpgalgetvsnvylgnastatkdkgnlkksdgtagnttttagatgtvngfagatahga 872  
QY 1922 VSVGASGAEERRIONVAAGEVSTATSDVANGSGLYKATQSIANATNELDHRITHONENKANA 1981  
Db 873 vsvgasgeerriqnvaaageisatstdaingsqlyavakgvcn---lagqvnkvvgkrada 928  
QY 1982 GISSAMAMSPQAVTIPGRSMVTGCIATHNGOGAVAGVLSKLSNDNGQWVVFKINGSADTQG 2041  
Db 929 gtsalaaasqlpqasmpgkmsvmsiagssyqgsglaigvrsldngkviirlsgttnsqg 988  
QY 2042 HVGAAGVAGGFHF 2053  
Db 989 ktgvaagvgvq 1000  
RESULT 12  
ID AAB23854  
X AB23854 standard; Protein; 1002 AA.  
XX  
AC AAB23854;  
XX  
XX 17-JAN-2001 (first entry)  
DT  
DE Haemophilus influenzae adhesin (Hia) protein from NTHi strain 33.  
XX  
XX Hia; adhesin; Haemophilus influenzae adhesin; NTHi; Infection; vaccine;  
KW

KW non-typeable Haemophilus influenzae; antinflammatory; auditory;  
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;  
KW diagnosis; immunogenic; antigen.  
XX  
OS Haemophilus influenzae.  
XX  
PN WO200055191-A2.  
XX  
PD 21-SEP-2000.  
XX  
XX 16-MAR-2000; 2000WO-CA00289.  
PF  
XX  
XX 16-MAR-1999; 99US-0268347.  
PR  
XX  
PA (CONN-) CONNAUGHT LAB LTD.  
XX  
XX Loosmore SM, Yang Y, Klein MH;  
PI  
XX WPI; 2000-618897/59.  
DR  
XX N-PSDB; AAA92493.  
XX  
XX Novel nucleic acid encoding Haemophilus influenzae adhesin protein, for  
PT use as antigens and vaccines and for treating Haemophilus influenzae  
PT infection  
XX  
XX Claim 1; Fig 18; 275pp; English.  
PS  
XX  
XX The present sequence represents a Haemophilus influenzae adhesin (Hia)  
CC protein from the non-typeable Haemophilus influenzae (NTHi) strain 33.  
CC Hia genes and proteins have antinflammatory, auditory and antibacterial  
CC activities, and can be used in the production of a vaccine. An  
CC immunogenic composition comprising an Hia gene, a polypeptide encoded  
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
CC protection against disease caused by Haemophilus strains in a  
CC susceptible host, preferably a human. An Hia protein is useful as an  
CC antigen, in immunogenic preparations including vaccines, as a carrier  
CC for other immunogens, and in the generation of diagnostic reagents. Hia  
CC is useful for treating diseases caused by the infection of Haemophilus  
CC influenzae such as meningitis, epiglottitis, septicaemia and otitis  
CC media. Recombinant production of Hia favours high recovery of the  
CC protein compared to the low recovery of native protein from Haemophilus  
CC influenzae species. A truncated protein has a significantly higher  
CC amount of recovery than a full-length protein.  
XX  
XX Sequence 1002 AA;

Query Match 9.2%; Score 952; DB 21; Length 1002;  
Best Local Similarity 28.5%; Pred No. 2.2e-34;  
Matches 324; Conservative 138; Mismatches 319; Indels 356; Gaps 41;

QY 1084 TGTNGSLDKSKPHLSKDGINAGGKKITNTQSGEIAKNS-----HDVATGCGKIYDLKTE 1136  
Db 55 tgn-----slnvygknnsfn-----annsladlnkgnsv-----ydglln 93  
QY 1137 LENKISSTAKTAQNSLHESVADQ-----GN-----NFTVSNPYSSYDTS---KTSVDIVF 1185  
Db 94 lnekgttdsk-----flvadettatvgnlrklgwwvstknstkeesngvqkadevlf 145  
QY 1186 AGENG---ITTKVNKGVVVRVIGIDOTKGLTTPKLTVGNN-----NGKGIVI----- 1227  
Db 146 egkgdgvvtvksengkhtvtfalandlnvknatvdsdklslgangkvvkdtstdangikfak 205  
QY 1228 ---NSQNGONTITGLSNLTANVNDKGSVFTTEGNTIREDKTRAASIVDVLISAGNLQ 1284  
Db 206 qgtngqngvnhlgiastldldprvggkthltke---isdternraasvqdvlnagwnlr 262  
QY 1285 -----GNGEAVDFVSTYDVFNFANGNNTTAKVTDYDTSKTSKVYDVVNNVDDTTIEVKDK 1338  
Db 263 gaktigtvcdnvdvstvydtvefasganansvttddnkktt-----vrvdvtgipvq-- 315  
QY 1339 KLGKVTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNTLSGDIQTAKGASQANNSA 1398

Db 316 ----- 315  
 Qy 1399 GYVDADGNKVIYDSTDNKYQAKNDGTVDKTEKAVAKDLV---AQAQTPDGTQLAQMNVKS 1455  
 Db 316 -yvtedskttvv---kvgnneyeakqgdsmkdkvengklaktkvklsangt-npvklsn 371  
 Qy 1456 VINKEQVNDV-----NKKOGINEDNAFVKGLEKAAASDNKTKNAATVVG-DLNVA 1504  
 Db 372 vadgtdtdavsfkqlkalqdkgtlslasnyang-----gsdadgkatqtlgndlnfkf 427  
 Qy 1505 QTP-----LTFAGDTGTAKKLGET-----LTIKGGQDITNKLTD-----NNIG 1543  
 Db 428 kstsdellnikaaagdtvtftpkksqvqvgddgkatidgakttgilveaselsvdslnklg 487  
 Qy 1544 VVAG-----TDC-----FTVKLAKDLTNLSNV- 1565  
 Db 488 wkvgvgdgtgatdgttdtvlvsgdkvtlkgdnlkvkqegtnftvyvlrdeltgkvksve 547  
 Qy 1566 -----NAGGTKIDEKISFV---DANGOA---KANTPVLISANGLDLGGKVISNVGKGT 1612  
 Db 548 fkdtdangastkitdglittpandangaaatdadkikvasdgisagnkavknvsvgl 607  
 Qy 1613 KDTDAANVQQLNEVRNLGLGNDAGNQVNIADIKDPNSGSSNRTVIRKAGTVLGGK 1672  
 Db 608 kkgdanfnpltsadnltkydnaykgltnldekskqkqtpvtvadnt---aatv----- 659  
 Qy 1673 NNDTEKLATGVGVQVDKGNANGDLSNV-WV---KTQKDGSKKALLATYNAAGQTNVYTN 1729  
 Db 660 -----gdrlglwvisadkktgeskeysaqvrnanevkfksg 696  
 Qy 1730 NPAAEIDRINQGIIRFFHVNDGNEPVPVQGRNGIDSSASGKHSVAIGFOAKDGEAAVAI 1789  
 Db 697 nglnvsgktldngtr-----eitfelakdena----- 723  
 Qy 1790 GROTQAGNQSIAGIDNAQATGDSIAIGTGNVAGKHSAGIDPSTV---KADNSYSVGN 1847  
 Db 724 -----lafsgskalrdntvaigtgnvnaeksgafgdpnyiedkaggsyafond 773  
 Qy 1848 NOFTDATQDVFVGNN-----ITV-----TESNSVALG 1876  
 Db 774 nrit-skntfvlngvnakykangdvdtetvtvkdkdgtttvtpkalgatvensvylg 832  
 Qy 1877 SNSAISAGTHAGTQAKKSDGTAGTTTAGATGTGKFPAGQTAGVAGSVGASGASERRIONV 1936  
 Db 833 nk-----statkdgnkksdgcagntttagttgvtngfagatahdavsvgasgeerriqnv 889  
 Qy 1937 AAGEVSATSDAVNGSOLYKATQSIANATNELDRIHONENKANAGISSAMAMASMPQAY 1996  
 Db 890 aageisatstdaingsqlayavakvtn---laggvnkvgradagtasalaasqlpqas 945  
 Qy 1997 IPGRSMVTGGIATHNGQAVAVGSLKSDNGQWFKINGSDATQGHVGAAGVAGPHF 2053  
 Db 946 msgksmvsiaqsgyqgqglagvrsisdngkvliirlagtttnsgktgvaagvgvqw 1002

RESULT 13

AAW56322  
 ID AAW56322 standard; Protein; 2039 AA.  
 AC  
 XX  
 AC AAW56322;  
 XX  
 DT 19-AUG-1998 (first entry)  
 XX  
 DE Haemophilus paragallinarum antigenic protein #2.  
 XX  
 KW Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;  
 XX vaccine; chicken infectious coryza; CIC; fowl.  
 XX  
 OS Haemophilus paragallinarum.  
 XX  
 FH key Location/Qualifiers  
 FT Peptide 1..70  
 FT /label= signal

FT Protein 71..2039  
 XX /note= "antigenic protein"  
 PN WO9812331-A1.  
 XX  
 PD 26-MAR-1998.  
 XX  
 PF 12-SEP-1997; 97MO-JP03222.  
 XX  
 PR 19-SEP-1996; 96JP-0271408.  
 XX  
 PA (KAGA ) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.  
 PA (KAGA ) CHEMO-SERO-THERAPEUTIC RES INST.  
 XX  
 PI Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;  
 DR WPI; 1998-230318/20.  
 DR N-PSDB; NAW22837.  
 XX  
 PT Antigenic polypeptide from Haemophilus paragallinarum induces HI  
 PT antibody production - and is useful for diagnosis of and preparation  
 PS of vaccines for chicken infectious coryza  
 XX  
 PS Claim 5; Page 71-87; 108pp; Japanese.  
 XX  
 CC The present sequence represents an antigenic protein derived from  
 CC Haemophilus paragallinarum strain C-53-47. The antigenic protein  
 CC stimulates the production of HI antibodies in fowl. The protein  
 CC and DNA coding for it can be used in the preparation of vaccines  
 CC for the prevention of chicken infectious coryza (CIC). The protein  
 CC and its antibodies can be used in the diagnosis and treatment of CIC.  
 XX  
 SO Sequence 2039 AA;

Query Match 7.8%; Score 809; DB 19; Length 2039;  
 Best Local Similarity 20.3%; Pred. No. 1e-27;  
 Matches 501; Conservative 311; Mismatches 810; Indels 848; Gaps 101;  
 Qy 1 MNHIYKVFNFNATGTFMAVAVKSHSTGGSCATCGQSVRTLSFARIAALAVLV---I 57  
 Db 1 mnkvfklysvvqemivvselannkdktsqknthttaffql-ftktylalllnial 59  
 Qy 58 GATL-----NGSAYAQQITTKIEIGQTNKIN-----NTLKGDMAL----- 91  
 Db 60 gtslfpqlanakfllevynssvklqhvnsqpsdsvn-lmpsggenvgmnsngvairga 118  
 Qy 92 -----ATGEASTAFGSLSKAOGSQAIAIGS----- 116  
 Db 119 vnnysatgsiaigqgakndnwafrsiaigqgaknesiasdsvalsnalnrfkksivigl 178  
 Qy 117 --VKPDPNNGSNGN-----VGSNAK--GNE-----SIAIGDVLAEGDASTA 154  
 Db 179 tytqlprapesrgsvvigenaksgnqsvslgnawsktnsisagagtfaegkstia 238  
 Qy 155 IGSDDLX-----LPAK--LDLKNFEH--KLINGHEILK----- 184  
 Db 239 igtdkilgntynyndklpapswdgirtgkaptnsiwdifselymgkngtdydkandrpn 298  
 Qy 185 ----IQTSTDGKIKY-----RRTRAQGHASTAVGAMSVAQGHFSN 220  
 Db 299 kpeafytsdfksryvnnpstptyaaklgalaalsrtlaagemstavgslafaladkst 358  
 Qy 221 AFG--TYATAEAAYSLAVGLAAQATKQSSIAVGSNAKAN-----AFAATAIGGNTVVN 271  
 Db 359 amglrfsfvakdvagvtaigeesrtfakdsvaigntkeasnaagsmaygykakavagagaia- 417  
 Qy 272 LGRGVALG-----FGSOILDRDNNDASAVVPLGKTLADQYKATROGSDSTDF- 319  
 Db 418 igaevaagaefdsqagnlllnr-----gayatlks--akdsddlikagdaivntqffd 469  
 Qy 320 -----SIGNSNNN-----NSSIRRKIIINVGAS--- 342

Db 470 nmltgshlytyentyltsagdiakktlaavdgdknaiaignkftaskansvalgsyal 529  
Qy 343 -RDTDAVNAQLKLVEL-ANKRITFKDGDNNNSVERG-----LGNTITRIK 388  
Db 530 asagnafalgysilvaplantivlgvyagysfnsvsgswstlsartvlgysasis 589  
Qy 389 GDAQTALTEANTGVYTDGNGLVKAKELTGLTSVATNKITVNTNNNNRELQSGGLT 448  
Db 590 sdshdsiamgvnafi--gng-----snssialg--- 615  
Qy 449 FSPITGKTDKTVYSIDGLKFTNDNSNIATKGTTRITKKKIGFAGTNDGV----- 498  
Db 616 ----tgsiaknakspdsialgkds-----ridakdldngvlytpqvydet 657  
Qy 499 -----DEKPYLDNPKLVKG-NSTLNSGSLTVNNTTKOIQVGANGIKFATVANNVA 550  
Db 658 trafrtdenkdy-m-rqamalgfnakvsrgkgme--tginsmaigar--sqatlgqnsa 712  
Qy 551 NTSATVGTARITEEKIGFAGTNDGV-----EQAPYLDKERL-----KVGREVEITDS 598  
Db 713 -----lgvnakdytweqleapdwskgaistskigvisv--- 750  
Qy 599 GINAGNKITGLTNGIANTDAVTKOLK----- 626  
Db 751 gskgserrlvnvasgsldtdavnavaqiktieerfgseidllqngggvqylsvsektninge 810  
Qy 627 -----DAKPTLUNAG--DGISIN-----SNNGD 646  
Db 811 agrvasqirgeskryvklktqilyldarkklngefdqtsidkiskavqealeaaysge 870  
Qy 647 LVDSSGNITTYNLSVKTTLNSNGTSGNNKFSVNAHDNNS-----LVTAKD 695  
Db 871 lkttaselnrvamqlnaett-vnfdgfkngyktqienatnadseknvgglsppqviaqlka 929  
Qy 696 LADYLNKVNETADSLPCKVQNGDNS-----NAITVGKDTNGKTFNTLKLGENGVNI 750  
Db 930 nnylndgagqgdsiafagqaktsaannaglagkqalaigfqans-----saenaisi 981  
Qy 751 TTN-----RATGTVTFIDGOSNGLT-----PKLTVGSDTFNGNRLVIEOVPSADGNS-TKNI 801  
Db 982 gtnsdtsmtgvalg-----kgatvtagkpsiaigqdstvansaisrtssvmlngltfnn 1037  
Qy 802 IKGLSPTLPSIA-SPSGR-----NIALGN-----TIEEKDKNAAISIDDLVN-AGFNLKN 850  
Db 1038 fagspetlglvsgtagkerkivnvaagdisqtsteaingsqlyatnfmnlkagsvktn 1097  
Qy 851 GKDKDFVSTYDVFIDGNATTATVTVYDEANQTSKVAVDVNVEKTIETLGDNGCKQLGV 910  
Db 1098 fggnanlatdgtitftniggtgqdtlndain-----nvltklslsateeeevsvg 1148  
Qy 911 KTIKLTETSTNGNATTFSTDDHALVKASDIAGNLNTLAEETHTKGTANTALQTFVKK 970  
Db 1149 eav---ydalkgaktvsaeankgiglvdvvkans---pitvepstdnkktftvgl 1202  
Qy 971 VDENDKADDTNATVKGDK-----TSGKVTNTLLKKNGLDKDKDGTFTVGINTQSG 1025  
Db 1203 mkd---legvnsitfdksqgdlnqvtgrmssagltfkkq---dttngsttf---aedgl 1253  
Qy 1026 KAGDSTT-----LNNNGLSIKNTASNEQIQVGADGVKFAMVNGVVGAGIDGTPRI 1076  
Db 1254 ti-dstnsaqatnlkvsvrsgdfsvkn-----gsdesklastkls-igaenaehvev 1302  
Qy 1077 TRDEIGFTGNGSLDKSKPHLSKDGINAGK-----KITNTOSGEIAKNSHDAVTGGKI 1130  
Db 1303 tksgiaiakadnts-dksaitlagdaitlagnatgtaitkltvgadnitvnskdavnggql 1361  
Qy 1131 YDL-----KTELENKISSAKTAQNSLHEFSVADPQGNFTVSNPYSYDYS 1177  
Db 1362 rtlilgvdsgaklggttekttisaisdvkqaltatdaykadmknk-----lv 1409  
Qy 1178 KTSVIVTFPAGNGITTVK-NGVVRVGI-DOTKGLTT-----PKLTVGN 1220  
Db 1410 kltadglntstnidasvedngvvkftikdltgltklatiesinasqnlaggctvtvgge 1469

Qy 1221 NGKGIIVINSQNGQNTITGLSNTLIANTVNDKGSVRRTEQGNIIKDEKTRAASIVDVLSAG 1280  
Db 1470 t-egivltksgsgndrtlsisgagnaatdgikvsgvkagt---adtdavnkgl-dklfka 1525  
Qy 1281 FNLGNGEAVDFVSTYDVTNFPANGNTTAKVTVDDTSKTSKVYDVYVDDTTIEVK---D 1337  
Db 1526 ln-----dalgttdlavtknpqts---ifnplngtapttkfkdavdklktavntgvg 1574  
Qy 1338 KKLGVKTTTITSTGCTGANKFALSNQATGDALVKASDIAHLNLTLSGDIQPAKASOANNS 1397  
Db 1575 skvgilatgldgidagnk--isnvadgd-----isptsgdvvtlgr----- 1613  
Qy 1398 AGYVDADGNKVIYDSTONKYYQAKNDGTVDKTEVAKDKLVAQAOTPDGTLAOMNVKSVI 1457  
Db 1614 -----qlyalmqglrvyvgdevspstktgttapt----- 1642  
Qy 1458 NKEQVNDANKKOGINEDNAFVKLEAKASDNKTNAAVTVGDLNAVAQTPLTFAGDTGTT 1517  
Db 1643 -----sstqggattantaggvapagvatgdiapt----- 1672  
Qy 1518 AKKLGETLTIKGGQDTNKLTDNNIGVVAGPDGFTVKLAKDLTNLNSVNAGGKTIKDEGI 1577  
Db 1673 qpaipekt-----alvgdhlavplgg---slkghdhvkvkttisag-----ndvg 1716  
Qy 1578 SF-----VDANGQAKANTPVLNSANGLDLGGKVISNVGKGTOTDAANVQQLNEV 1626  
Db 1717 siqpnisiennlivigsnkpekaklaaqegnal-----vitn-----kddgnaamv--fne 1765  
Qy 1627 RNLGLGNDNADGNQVNIADIKKDPNS---GSSNRRTVIKAGTVLGGKGNNDTEKLATGG 1683  
Db 1766 kmnlv-----sdkkakpravldgngaltlv-----gnddsq--vfls 1802  
Qy 1684 VQGVVDKCGNANGDLSNVWVKTKDCKSKKALLATVNAAGQTVNNTNPAEALDRINEOGI 1743  
Db 1803 skkgkdidgn---dlarlsvttert-----nadgglekvetsfatmdd-----gl 1844  
Qy 1744 RFPHVNDGQEPVQGRNGIDSSASKHSVAIGFOAKKADGEAAVAIGRQTOAGNOSIAIG 1803  
Db 1845 kf-----kadgdkvi-----nkkln 1859  
Qy 1804 DNAQATGDGSIATGTCNVVAGHSGAIGDPSTVKADNSVSGVNNQFTDATOTDVFVG 1863  
Db 1860 etvelvgdenvt-----tsitddnkvkvslnkkla-----idevklpn 1897  
Qy 1864 NIIVTESNVALGSNSAISAAGTHAGTQAKKSDCTAGTTTTAGATGTVKGFACQAVGAVS 1923  
Db 1898 tdpdaqgdsdivlnnggihagnkvilgvtkasd----- 1929  
Qy 1924 VGASGAERRIQNVAAAGEVSATSTDAVNGSQLYKATOSIANATNELDHRHONENKANAGI 1983  
Db 1930 -----dptsavnrqglntvldnvgnnfnqvnqrfgdltresragl 1969  
Qy 1984 SSAMAMASMPQAVIPGRSMVTGGIATHNGGAVAGVGLSKLSDNGQWVFKINGSADTQGHV 2043  
Db 1970 agemataslgnvalpgkttisvgtatfkgenavaimrslsdngkvgrlrgsmatsngdk 2029  
Qy 2044 GAAVAGAGFHF 2053  
Db 2030 gaamsvgftf 2039

## RESULT 14

AAW56319

ID AAW56319 standard; Protein; 2042 AA.

XX AAW56319;

AC AAW56319;

XX 19-AUG-1998 (first entry)

XX Haemophilus paragallinarum antigenic protein #1.

XX Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;

KW

vacine; chicken infectious coryza; CIC; fowl.

KW XX OS XX FH XX FT XX FT XX FT XX PN XX XX XX PD XX XX PF XX XX PR XX PA XX FA XX PI XX DR XX XX PT XX PT XX PS XX CC XX CC XX CC XX CC XX CC XX SQ

Haemophilus paragallinarum.  
Key Location/Qualifiers  
Peptide 1..70  
Protein /label= signal  
71..2042  
/note= "antigenic protein"  
W09812331-A1.  
26-MAR-1998.  
12-SEP-1997; 97WO-JP03222.  
19-SEP-1996; 96JP-0271408.  
(KAGA ) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.  
(KAGA ) CHEMO-SERO-THERAPEUTIC RES INST.  
Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;  
WPI: 1998-230318/20.  
N-PSDB: AAV22834.  
Antigenic polypeptide from Haemophilus paragallinarum induces HI  
antibody production - and is useful for diagnosis of and preparation  
of vaccines for chicken infectious coryza  
Claim 1; Page 51-69; 108pp; Japanese.  
The present sequence represents an antigenic protein derived from  
Haemophilus paragallinarum strain A-221. The antigenic protein  
stimulates the production of HI antibodies in fowl. The protein  
and DNA coding for it can be used in the preparation of vaccines  
for the prevention of chicken infectious coryza (CIC). The protein  
and its antibodies can be used in the diagnosis and treatment of CIC.  
Sequence 2042 AA;  
Query Match 7.5%; Score 779.5; DB 19; Length 2042;  
Best Local Similarity 20.4%; Pred. No. 2e-26;  
Matches 506; Conservative 316; Mismatches 790; Indels 871; Gaps 111;  
QY 1 MNHIVKVFNNKATGTFMAVAEYAKSHSTGGSCATQGVSVRTLSFARIAALVLV---I 57  
DB 1 mnkvfkikysvvqemivvslannkktasqkthntaffqpl-ftkctylallinial 59  
QY 58 GATL-----NG---SAYAQQTITKIEIGQTNKIN-----NTLKG-----D 89  
DB 60 gaslfpqlanakwilevssvsklsvsagsnvnlnpspsvgtntspgvaigyatnd 119  
QY 90 ALATGEASIAFGSLSKAGSQAIAIG-----SVKP-----119  
DB 120 rsagalaigvgvkneltakdisiaigyaknestapstvigkqainfeksivmglmay 179  
QY 120 ---DPNNGSNGN-----VGSNAK--GNE-----STAIGGDVLAEGDASTAI 155  
DB 180 tqldprglsketrqgsvvigenaknsagqsvslqgnsksktnsisagatfaegkssiai 239  
QY 156 GSDDL-----YLPKN--LDLKNFPH--KLHGHEILKIKIQTST-----189  
DB 240 gtdklsigtkyndkipatawngtgtvpknslwdfiselymgkqtngrd---ydttrdqn 295  
QY 190 -----DGKIKY-----RRTPRAOCHASTAVGAMSYAGQHFSN 220  
DB 296 kpeafykfsdkgkyvntptasptyagklgaialgrstiaaagemstavgslafaladrst 355  
QY 221 AFG---TYATABAAYSLAVGAQAATKQSSIAVGSNAKAN-----AFAATAIGGNTVVN 271  
DB 356 amglrsfvakdavggtalgeesrftfakdsvaigknkneasnagsmaygykakavagagaia- 414

QY 272 LGRGVALG--FGSQ---ILDRDNNTDASAYVPLGKTKLADQYKATRQGDSTDFISGNSN- 325  
DB 415 igtevaagakfnshqgtgnllqdn---ayatlkn--adksddtktnaitvftgsfdnm 468  
QY 326 -----NNSSIRRKIINVGAGSRDQDANVA-----351  
DB 469 lnglplvsenetyltsagalkttattdsagggknalaigsktfaskansvalasyal 528  
QY 352 -----QLKLVEELANRKITF-KGDDNNSNSVBERG-----LGNLTIRIK 388  
DB 529 adaqnaafalgysfvessatntitigvsgyakgnslfggtwastlstrtvvlgnsstis 588  
QY 389 GDAQTNALTEANIGV-VTDGNGLKVKLAKELTGLTSVSATN-----KITVSNTNN 437  
DB 589 sgsg-nala---igvvnfigndsasslalmgngstiaaksakpsdslaigkearidaktdn 644  
QY 438 NNAELQSGGLTFSPITGKTKDVTYVSDGLKFTNDNSNIATKGTTRTKKIKGIFAGTNDG 497  
DB 645 -----gtlygpqvdy-----ettraf-----n 662  
QY 498 VDESKPYLDNEKLKVGNSLTNLSGSLTVNNTTGNKQIOVCANGIKFATVANNVANTSATVG 557  
DB 663 fnessdym-rqamalg-----fnakvsrgvgkmetginsmaigayaqatlnstalg 713  
QY 558 TARITE---EKIGFAGTNDGVEQAPYLDKERLKVGRVEITFDS-----GINAGNHKI 607  
DB 714 vgsktdytwegl-----etdpwvse-----gaisiptsgktvisvsgksgerri 758  
QY 608 TGLNGIANTDAVTIKQLKDAKP-----TLNAGDGI-----SINSN-----643  
DB 759 vnlasgssdtadvnvaqiktveerflselnllqngggvkyisevktningdsggrvasqir 818  
QY 644 -----NGDLVDS-----SGNITTTTYNIS 662  
DB 819 kgenyeryvkkltqlyldargklngkefkdnslkniravvqeleaaesgelkttasaln 878  
QY 663 VKTKLNSGSGN-NKFS-----VSNADHNSL-----VTAKDIA-----DYLNKYN 705  
DB 879 qvatqleqvettnnfdknqyktqienasnasdsarnvvgltppqalqakannnyindgak 938  
QY 706 TADSLALPFKQVQ-----NGDSSNNAITVGDKNKGTFTWTLKLGKGVNITN---RATG 757  
DB 939 qgdsiafvgqaktsgannglagkaiaifqans-----saenaislgnsdtsmctg 990  
QY 758 TWTFGIDSGNLT-----PKLTGSDTNGNRLVIEQVPSADGNSTKNIKGL-----805  
DB 991 avaiq---kgatvtagkpsialqgdst-----vansaisrtsspminglifnftag 1039  
QY 806 SPTLPASIAS--PSGR-----NIALGNTIEEKDKS-----NAASIDDLVNA 843  
DB 1040 spetlgvisigtagerkivnvaagdvsgasteaingsqlyatnfmkskvagsvksnfgg 1099  
QY 844 GFNLKNNCKDXDFVSTYDTPIDGNATATVYDEANOTSKVAYDVNVNDEKTIETGDN 903  
DB 1100 nvnlgtdg-----tftfnl99tgqatlihaainvlckgylkadtgn--dptngq 1147  
QY 904 GKK-OLGVKTKIKLTET---STNG-----NATFTSDDDHALVKASD-----TAGNLN 946  
DB 1148 gqkelvg-naitlsatnqwanngvnyktnnltynsqngtllfmgredpsvkqitagtyn 1206  
QY 947 TLAEEIHTTKGTANTALQTFVTKVVDENKXADDTN--AITVGKDG---TSGKWNITLKG 1001  
DB 1207 ttgda--nnknqnlhntlqgttleatgitssvgsstnyagfslgadsvtfskkgagvklsg 1264  
QY 1002 KNGLDKTKDKG-----TVTFGTNGTOSGLKAGDSTLLNNGLSI-----KNTASN 1046  
DB 1265 vs--dataatdaatlkvkeyrtllvgdn---ditaadrsrgstngitnynlslnkgtvsa 1319  
QY 1047 EQIQVAGDGVKFAFMYNNGVVG-----AGIDGTTTRITRDEIGFTGTNGSLDKSKPHISKDGI 1102  
DB 1320 teekvsvgktyeairnaitgnifiglddt-----lnkinnpadqdlslnsesgk 1371

1103 NAGGKITNLOSSEIAKNSHDAVTGKIIYDLKTELENKISSTAKT-AQNSLHFSVADEQ 1161  
1162 GNNFTVSNPYSSYDTSK---TSDVITFAGENGITTKVNGVVRVIGDITQGLITPKLTV 1217  
1422 --kltksvesytnklnfstldillsdgrsgnattandvgkrlsdgfftkksentfl 1479  
1218 GNN--NGK---GIVINSQNG---QNTITGLSNTLANVTNDKGSVRTTEOGNIKDEDKT 1268  
1480 gskdyngdsalgvmyddqngvfxlsnmtalttslan-----tfakidasnltdsne 1533  
1269 RAASIVDVLGAGENLQNGEAVDFSVSYDTVNFANGNTTTAKVYVDDTSTKTSKVYDVNV 1328  
1534 kwtalnvs-----ktevdaelqkskvtl---t 1559  
1329 DDTTIEVKDKKLGVKTTTLSTGTGANKFALSNOQTGDALVKASDIVAHLNLSGDIOFA 1388  
1560 pdsqlifatkdag---sgnnagidagnkkslnvadg-----isptsgdvvtg 1604  
1389 KGAOANNSAGYVDADGNKVIYDSTDNKYQAQNDGTVDKTEKVKDKLVAQAQT---PD 1445  
1605 r-----qlvalmqkgirvvygdevsptkqtaptanpt 1638  
1446 GTLAQMNVKSVINKEQVNDANKOGINEDNAFVKGLEKAASDNKTKNAATVVGDLNAVAQ 1505  
1639 al-----taptasstgqwaltantagvav 1663  
1506 TPLTFAGDTGTAKKLGETLTIKGQDTNKLTDNNIGVAVGDTGTVKLAKD---LTNL 1562  
1664 agnavtdgiaptcptlpemt-----alvdhlaavlpgg---sikhghdhvkttl 1711  
1563 NSVNAAGGTIDKSGISF-----VDANGAKANTPVLISANGLDLGGKVISNKGTKDIDA 1617  
1712 sadnqvglsl-qpnisiennlviqsdpekaklaaqegnal-----vitn-----kddgna 1761  
1618 ANVOQLNEVRNLLGLGNDNADGNQVNIADIKKDPNSGSSNRVTWIKAGTVLGKG----- 1671  
1762 amv--fnneknmvltsdkea-----kprvllldgngaltl 1794  
1672 -GNNDTEKLATGGVQGVVDKNGDLSNVWVKTKDGSKKALLATYNAAGQTNVYTN 1730  
1795 vgnddsg--vtlsskkgkldgn---dlrlsavttert-----nadgglekvets 1839  
1731 PARAIIDRINQGTREFHVNDRQEPVQVQNGRIDSSAGKHSVAICGFAKAGAAVAIG 1790  
1840 fatmdd-----gkf-----kadgkvi--- 1857  
1791 ROTOAGNOSIAIGDQAQATGDSIAIGTGNVAGKHSAGIDPSTVKADNSVSGNNQOF 1850  
1858 -----nkklnetvelvgdenvt-----tsltdonkvkslnkki 1891  
1851 TDATQDVFVGNNTVITNESNVALGNSAISAGTHAGTQAKKSDGTAGTTTAGATGTV 1910  
1892 a-----idevkipntdpdaqkdsivimnggihagnkvilgvkasd----- 1932  
1911 KGFAGQTAGVAVSVGASGAERRIQNVAAGEVSATSDAVNGSOLYKATOSIANATNLDH 1970  
1933 -----dptsavnrqglntvidnvgnnfnqng 1959  
1971 RIHONENKANAGISSAMAMSPQAVYIPGRSMVTGGIATHNGOCVAVAGVLSKLSNDGQWV 2030  
1960 rigdlrtresragiagataslqvalpgkttisvgtatfkgenavagmsrlsdngkv 2019  
2031 FKINGSADTQGHVGAAGVAGPHF 2053  
2020 irlsgmstsnkgdaemsvgfsf 2042

RESULT 15  
AAB23858  
ID AAB23858 standard; Protein; 1094 AA.  
XX

AAB23858;  
17-JAN-2001 (first entry)  
Haemophilus influenzae adhesin (Hia) protein from NTH1 strain K9.  
Hia; adhesin; Haemophilus influenzae adhesin; NTH1; infection; vaccine;  
non-typeable Haemophilus influenzae; antiinflammatory; auditory;  
antibacterial; meningitis; epiglottitis; septicaemia; otitis media;  
diagnosis; immunogenic; antigen.  
Haemophilus influenzae.  
WO200055191-A2.  
21-SEP-2000.  
16-MAR-2000; 2000WO-CA00289.  
16-MAR-1999; 99US-0268347.  
(CONN-) CONNAUGHT LAB LTD.  
Loosmore SM, Yang Y, Klein MH;  
WPI; 2000-618897/59.  
N-PSDB; AAA92497.  
Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for  
use as antigens and vaccines and for treating Hemophilus influenzae  
infection  
Claim 1; Fig 22; 275pp; English.  
The present sequence represents a Haemophilus influenzae adhesin (Hia)  
protein from the non-typeable Haemophilus influenzae (NTH1) strain K9.  
Hia genes and proteins have antiinflammatory, auditory and antibacterial  
activities, and can be used in the production of a vaccine. An  
immunogenic composition comprising an Hia gene, a polypeptide encoded  
by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
protection against disease caused by Haemophilus strains in a  
susceptible host, preferably a human. An Hia protein is useful as an  
antigen, in immunogenic preparations including vaccines, as a carrier  
for other immunogens, and in the generation of diagnostic reagents. Hia  
is useful for treating diseases caused by the infection of Haemophilus  
influenzae such as meningitis, epiglottitis, septicaemia and otitis  
media. Recombinant production of Hia favours high recovery of the  
protein compared to the low recovery of native protein from Haemophilus  
influenzae species. A truncated protein has a significantly higher  
amount of recovery than a full-length protein.  
Sequence 1094 AA;  
Query Match 6.9%; Score 710; DB 21; Length 1094;  
Best Local Similarity 25.1%; Pred. No. 1.1e-23;  
Matches 315; Conservative 162; Mismatches 438; Indels 340; Gaps 56;  
QY 953 HTTKGTANTA---LQTFVKKVDENDKADDTNAITVGKDTGKVTNKLKGNGL-DIK 1008  
Db 26 htkrasatvatavlatqlsatacaanssasytsrlnv-----ygdtnkfknaannsladln 80  
QY 1009 TDKGDTVFGINTOSGLKAGDSTLTNNNGLSIKNTASNEIQVAGDGVKRFAMVNNGVGA 1068  
Db 81 kqndg-----vhdgl-----lnhngankllvdntaalvgdlrklgvvstknkg 128  
QY 1069 GIDGTTTRTR-DEIGFTGTNGSLDKSPHLSKDGINAGGKKTNIQSGEIAKNSHDAVTC 1127  
Db 129 eneksqqvkqadevlfksgk-----gvqvtstseengkhaif- 165  
QY 1128 GKHYDLKTELENKISSTAKTQAQNSLHFSVADEQGNNTVTSNPNYSSYDTSKTSVITFAG 1187  
Db 166 ---falakaldmr-----tatvsdtlttgg 187

Qy	1188	ENGI	TTKVNKGVVRVIGIDTQGLTTPKLTIV	-----GNNNKGIVINSQNGONTI	-----TGUS	1244	
Db	188	s----	tt			229	
Qy	1241	NTLANVTNDKGSVRTTEOGNI	IKDBKTRAASIVDVLSAGFNLOQ	-----NCEAVDEFS	1294		
Db	230	stlqdtlllncgvvskld	-ngltdakkrasvqdvinsgnlkgvktgattsndvdfvr	288			
Qy	1295	TYDTVNFANGNTTTAKVYDDT	SKTSKVVYDVNVDDTTIEVKOKKLGKVTTLTSTGTG	-1353			
Db	289	tydtveflsgseettlvtvdsesng	kskvkigaktsvikekdgl	-----fgkan	340		
Qy	1354	--ANKFALSNOA	-----TGDAIVKASDIVAHLNLTLSGDIOTAKGASQANNSAGYVD	-----1402			
Db	341	kdlnqvasnaaaddtdedgklv	taetvlnavnkagwrkt	-----tganqagqfetvts	396		
Qy	1403	-----ADGN	-----KVLYDSTON	-----KYFOAKNDG	-----TVDRKTK	-EVAKDKLVAQ	1440
Db	397	tnvtfadngttavvtgdating	itvkyeakvgdglkigndqdkltadttaltvfgkvta	-455			
Qy	1441	AQPPDGT	LQAMNVKSVINKEQVNDANKKOGINEDNAFVKGLEKRAASDNKTKNAAVTVGD	1500			
Db	456	---pdat	-----ngkklvnasgladalnk	-----lswtakeaad	-cang	---gel	493
Qy	1501	NAVAQPTPLTFAGDTGTTAKK	LGETLITIKGGDTDTNKLTDNNGVVGAGTDFGTVKLAKDLT	1560			
Db	494	d-----	-----qtadekevketvtfkag	-----nlkvkdganftyslqdat	534		
Qy	1561	NLNSV	-----NAGGPKIDEKGISFVDANGQA	-----KANTPVL	SANGLDGLGKVKSVNYKGT	1612	
Db	535	gltsiltlgnngakteinkdgl	ttitpangagannantistvkdgisagqgskvnyvsgl	594			
Qy	1613	KDPTDAANVOQL	-----NEVRNLLGLGNDNDAGNOVNIA	-----1646			
Db	595	kkfgdanfpdltsaadnltk	gydddykgltnldekga	dkltlvadntaatvtdgl	-glgw	654	
Qy	1647	-IKKPNSG	-----SSSNRTV	IKAGTV	-LGGKGNNDTEKALTGGVOGVKDXDGNAGDL	-----1698	
Db	655	visadkttdgldekeynaqr	naevnkfksnginsvktvngrrrlfel	-----akgevsk	711		
Qy	1699	SNVYKTDQDKSKALLAT	-----YNAAGOTNTVTN	NPABADRINEQOIRFHFVNDGNQEP	1755		
Db	712	sneftvketngketslvk	vgdkyyskedi	patgkp	-----kvtgnnava	756	
Qy	1756	V-VQGRNGIDSSASGKHSVA	IGFOAKADGEAAVAIGROTQANGQSIAIGDNAATGDQSI	1814			
Db	757	akyqdkdkvksaadssntav	tltnkgygy	-----tgnqv	-----adaia	kskf	801
Qy	1815	AIGCTGNWAGKHS	GATIGDPS	-----TVKADNS	-----YSVGNNOQFTD	ATDTPDFCVGN	1863
Db	802	elglad	-aekakaafgdetkalsd	kdletvnan	dkvrfanglntkvsaaatvesidangd	859	
Qy	1864	NITVT	-ESNSVALGNSAISAGTHAGTQAK	-----KSDGTAGTT	-TTAGATGT	---1909	
Db	860	kvtttfvkdvelpt	qiyntdangkkivngdkwyytkdkdgdmdckmckvtlgnvdsdgk	919			
Qy	1910	-----VK	-CFAGQTA	V-----GAYSVG	-----ASCAERRIQ	NVAAGEV	1941
Db	920	kvvkednkw	hvksgdstdkttqvveeakvs	dekhhvslpnddskgkgyvinmangel	979		
Qy	1942	SATSTDAVNGSOLY	-----KATQSI	ANATNELDHRITQHONENKANAGISSAMAMASMPQAI	IP	1998	
Db	980	satsdaingsql	yavakvtnlaqgvnnlegkvkvgk	raadaqtasaalsqlpatamp	1039		
Qy	1999	GRSMVTGGIATINGOGA	VAVAGLSKUSDNQWFKI	NGSADTQGHVGAAGVAGPHF	2053		
Db	1040	gkmsvlsagsgvgngl	algvrsldnqgkvlrlsglttnsgsktgvaaqvgw	1094			



\_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:37:49 ; Search time 44.72 Seconds  
(without alignments)  
945.259 Million cell updates/sec

Title: US-09-361-619-9  
Perfect score: 10356  
Sequence: 1 MHHIYKVIENKATGTFMAVA.....NGSADTQGHVGAAGAGHFH 2053

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7357	71.0	2123	4	US-08-968-685A-10
2	1206	11.6	2353	4	US-09-377-155-33
3	1206	11.6	2353	4	US-08-913-942-4
4	890	8.6	1912	1	US-08-409-995-4
5	890	8.6	1912	3	US-08-685-457-4
6	690	6.7	1098	1	US-08-409-995-2
7	690	6.7	1098	3	US-08-685-467-2
8	690	6.7	1098	4	US-09-377-155-32
9	690	6.7	1098	4	US-08-913-942-2
10	583.5	5.6	679	4	US-08-913-942-15
11	546.5	5.3	1600	2	US-08-617-697-10
12	515	5.0	1599	2	US-08-617-697-9
13	512	4.9	1529	2	US-08-728-470-10
14	512	4.9	1529	4	US-08-719-641-10
15	469	4.5	1477	1	US-08-038-682-4
16	469	4.5	1477	1	US-08-302-832-4
17	469	4.5	1477	2	US-08-530-198-4
18	469	4.5	1477	2	US-08-469-880-4
19	469	4.5	1477	2	US-08-728-470-4
20	469	4.5	1477	2	US-08-617-697-4
21	469	4.5	1477	4	US-08-719-641-4
22	459	4.4	1338	2	US-08-728-470-9
23	459	4.4	1338	4	US-08-719-641-9
24	456	4.4	1536	1	US-08-038-682-2
25	456	4.4	1536	1	US-08-302-832-2
26	456	4.4	1536	2	US-08-530-198-2
27	456	4.4	1536	2	US-08-469-880-2

28	456	4.4	1536	2	US-08-728-470-2	Sequence 2, Appli
29	456	4.4	1536	2	US-08-617-697-2	Sequence 2, Appli
30	456	4.4	1536	4	US-08-719-641-2	Sequence 2, Appli
31	440	4.2	1612	1	US-08-169-927-2	Sequence 2, Appli
32	393	3.8	599	4	US-09-377-155-15	Sequence 15, Appli
33	391	3.8	592	4	US-09-377-155-17	Sequence 17, Appli
34	388	3.7	594	4	US-09-377-155-9	Sequence 9, Appli
35	387	3.7	598	4	US-09-377-155-5	Sequence 5, Appli
36	385.5	3.7	592	4	US-09-377-155-2	Sequence 2, Appli
37	385	3.7	598	4	US-09-377-155-13	Sequence 13, Appli
38	384.5	3.7	658	1	US-08-409-995-5	Sequence 5, Appli
39	384.5	3.7	658	3	US-08-685-467-5	Sequence 5, Appli
40	384.5	3.7	658	4	US-08-913-942-5	Sequence 5, Appli
41	382.5	3.7	589	4	US-09-377-155-19	Sequence 19, Appli
42	381.5	3.7	591	4	US-09-377-155-11	Sequence 11, Appli
43	381.5	3.7	591	4	US-09-377-155-21	Sequence 21, Appli
44	377	3.6	594	4	US-09-377-155-7	Sequence 7, Appli
45	353	3.4	607	1	US-08-409-995-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-08-968-685A-10  
; Sequence 10, Application US/08968685A  
; Patent No. 6214981  
; GENERAL INFORMATION:  
; APPLICANT: TUCKER, KENNETH  
; APPLICANT: PLOSILA, LAURA  
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/968 685A  
; FILING DATE: No. 6214981ember 12, 1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baldwin, Geraldine F.  
; REGISTRATION NUMBER: 31,232  
; REFERENCE/DOCKET NUMBER: 7969-060  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2123 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-968-685A-10  
  
Query Match 71.0%; Score 7357; DB 4; Length 2123;  
Best Local Similarity 70.6%; Pred. No. 0;  
Matches 1571; Conservative 120; Mismatches 261; Indels 272; Gaps 34;  
  
QY 1 MHHIYKVIENKATGTFMAVAEYAKSHSTGGSCATCGVSVRTLSFARIAALAVIGAT 60  
|||||

Db 1 MNHUYKVENKATGTFMAVEYAKSHSTGGSCATGVGVSVRTLSFARIAALAVLVIGAT 60  
QY 61 LNSAYAOQITTKIEIGQTNKLNNTLKGDAATGEASIAFGLSKAOGSOAIAIGSVKPD 120  
Db 61 LNSAYAGISEADGG-----KGANARGDKSIAIGDIAQLGQSIAIGDNKIV 111  
QY 121 PNGSNGVSHAKGNESIAIGDVLAECDASTAIGSDDLPLPK-NLDLKNFHKHLHGH 179  
Db 112 HSNNNANIGAKAGNESIAIGDVLASGHASIAIGSDDLYLKKTQQOISELLPIRQG 171  
QY 180 EILKIKIOTSDGI-KYRTRAGHASTAVGAMSTAQGHFSNAFGTYATAEAYSLAVGL 238  
Db 172 KALNDIYQLADTNLQYRTHRAOCHASTAVGAMSYAKGHFSNAFGTRATAGTYSLAVGL 231  
QY 239 AAOATKOSSTAVGASNAKANAFAATAGGTVNVLGRVALGFGSQTLDRDNNTDAS--A 295  
Db 232 TATAKAASSIAGVSNAGIAIGAATAVGGSTOVNLRGIALGFGSQVLQKDNVANAURA 291  
QY 296 YVP-LGKTLADQYKATROGSDSTDFISGNSNNNSIRKLIINVAGAGSRDTPAVNVAOLK 354  
Db 292 YAPDNNQPIDNRYKATFKNGATDFESGNSGND-SIRKLIINVAGAGSADTPAVNVAOLK 350  
QY 355 LVEELANRKITFKGDDGNNNSVERGINTLTIKGDAQTNALTEANIGVVVTDGNGLKVKL 414  
Db 351 EAVRLANROITFKGDDSN--NRVEKGLGKLTITVGGAQTSALTQHNIGVVVQNGDGLKVQL 408  
QY 415 AKELTGLTSVATNKTIVTSNTNNNAELQSGGLTFSPITGKTDKTVYSIDGLKFTNDSN 474  
Db 409 AETLSLKWVTEN-----LT-----AN 426  
QY 475 SIATAGTTRITKKGIPAGTNDGVDESKPYLD----- 506  
Db 427 EKVTVGKTRLTGKTGFTNDMNGIDESKPYLDKDTGTHAGGQKITKLTAGVWDDAATYG 486  
QY 507 ----- 506  
Db 487 QLKVNQTAESALOTFTVKVKDKNANDANSKIITVGNKNNKPDGTQVNTLKLGENGVGV 546  
QY 507 -----NEKLKVGNSLNSGLTVNNNTGNKQIOVGANGIKFATVANNVANT 552  
Db 547 TTETNGTVTFGLNONGNLTVGNSTLNNGLSVKNTNSNKQIOVGADGITTFDISNSKPG- 605  
QY 553 SATVGTARITEKIGFAGTNDGVDSQAPYLDKERLKVGRVEITTDGSGINAGNHKLTGL-- 610  
Db 606 AGENTTRITRDIAGFANNTGSLDANKPRL-----TPTGINAGGKELTNVQS 652  
QY 611 -----TNG-----TANTD-----AVTIKOLDAK--PTLNAAG--GISINSNGDLV 648  
Db 653 AINPATNGGQLDPMNRLSTANTKSGSAATIKDLNLSQVPLTFAGDTGPNVTKLGEI- 711  
QY 649 DSSGNITPTYNISVKTTLKNSNGTSNNKFSVSNADHNNLSV-TAKDLADYLNKVN-- 705  
Db 712 -----LKVGGKTADDLTKNIGVADSTNLSVLAKTLSL-LDAVNTKT 758  
QY 706 -----TADSLPSPKQVQNGSNNAITVGKDTNGKT--FNTLKLKNGGVNITNRA- 755  
Db 759 LTASDKVTVDGSGNNTAKLQND-----LTFSKONTGATPATNSKTIYGVGDKLFTDNGI 813  
QY 756 --TGTV-----TFGIDOSNG-----LTPKLTVGS--DTNGNRLVIEQVPSADGNS 797  
Db 814 ALDGTYYITKDKVGFAGKODGSLDKSRPYLDKDLKVGVEITNGI-----NA 861  
QY 798 TKNIKGLSPLTPIASPSGRNIALGNTIEEKSNASIDDLVNLNAGFNLKNGKDKDFV 857  
Db 862 GKATITGLSNTLTDATNATTGHVTLQGLVDSSTDKTRAASIGDVLNAGFNLKNGDKAKDFV 921  
QY 858 STYDTPDFIDGNATATVYDEANOTSKVAYDVNDEKTIETLC-DNGKKQLGKVKTKLT 916  
Db 922 STYDTPDFEINGNATKATYD--GKASKVAYDVNVDGTTIHLTGADGNKNOIGVKTTILT 979  
QY 917 ETSTNGN-ATTFSTD--DDHALVKASDIAGNLTLAEIHTTKGTANTALQTTFTVKKVD 973  
Db 980 KTDAGDKAINFVSNSGDDKALINAKDITADNLNLAGEIRNTKGTADTALQTTQVKKVKE 1039

QY 974 N-DKADDTNALTIVGKDTSGKYNTLKLKGNGLDIKTDKDTGTVTFGINTQSGLKAGDSTT 1032  
Db 1040 NGDDDDADADITIVGDKAKTNQVNTLKLKGNGLDIOTNKDGTGTVTFGINTQSGLKAGNNTT 1099  
QY 1033 LNNGLSIRKNTASNEOIQVAGDGVFAMVNVNGVYAGIDGTTRITRDEIGTGTNGSLDK 1092  
Db 1100 LNNGLSIRKNTAGNEOIQVAGDGVFAKVNNGVYAGIDGTTRITRDEIFAGTNGSLDK 1159  
QY 1093 SKPHLSKDGINAGGKKTITIOSGEIAKNSHDAVTGGKIYDLKTELENKISSTAKTAQNSL 1152  
Db 1160 SKPHLSKDGINAGGKKTITIOSGEIAQNSNDAVTGGKIYDLKTELENKISSTAKTAQNSL 1219  
QY 1153 HEFSVADQEGNNFTVSNPYSSYDTSKTSDVITFAGENGITTKVKNKCVVRVGIDQTKGLTT 1212  
Db 1220 HEFSVADQEGNNFTVSNPYSSYDTSKTSDVITFAGENGITTKVKNKGVVRVGIDQTKGLTT 1279  
QY 1213 PKLTVGNNGKGIIVNSONGONTITGLSNTLANVNDKGSVRTTEQGNIIKDBDKTRAAS 1272  
Db 1280 PKLTVGNNGKGIIVDSONGONTITGLSNTLANVNDKGSVRTTEQGIKIKEDDKTRAAS 1339  
QY 1273 IVDVLSAGFNLOQNGEAVDFVSTYTYNFANGNTTTAKVYDDTSTKTSKVYDVVYDDTT 1332  
Db 1340 IVDVLSAGFNLOQNGEAVDFVSTYTYNFADGNATTAKVYDDTSTKTSKVYDVVYDDTT 1399  
QY 1333 TEVKDKKLGVWTTTLTSTGTGANKFALSNOATGDALVKAADIVAHLNLTSCDIQTAKGAS 1392  
Db 1400 TEVKDKKLGVKTTTLTSTGTGANKFALSNOATGDALVKAADIVAHLNLTSCDIQTAKGAS 1459  
QY 1393 QANNSAGVVDADGNKVIYDSTDNKKYQAKNDGTVDKTEKAKDKLVAQAQTPDGTLAQMN 1452  
Db 1460 QANSSAGVVDADGNKVIYDSTDNKKYQAKNDGTVDKTEKAKDKLVAQAQTPDGTLAQMN 1519  
QY 1453 YKSVINKEQVNDANKKQGINEDNFAFKLEKASDNKTNAAVTVGDLNAVAQTPLTFAAG 1512  
Db 1520 YKSVINKEQVNDANKKQGINEDNFAFKLEKASDNKTNAAVTVGDLNAVAQTPLTFAAG 1579  
QY 1513 DTGTTAKKLGETLTITKGGQTDNTKLTNNIGVWAGTDTGTVYKLAKDLTNLNSVWAGTKI 1572  
Db 1580 DTGTTAKKLGETLTITKGGQTDNTKLTNNIGVWAGTDTGTVYKLAKDLTNLNSVWAGTKI 1639  
QY 1573 DEKGSFVDANQOAKANTPVLSSANGLDLGGKVISNVGKGTGKTDAAVQOOLNEVRNLGL 1632  
Db 1640 DDKGVSFVDSGQOAKANTPVLSSANGLDLGGKVISNVGKGTGKTDAAVQOOLNEVRNLGL 1699  
QY 1633 ---GNDNADGNQVNIADIKKDPNSGSSSNNRTVIKAGTVLGGKGNNDTEKLTGQVGVGD 1689  
Db 1700 GNAGNDNADGNQVNIADIKKDPNSGSSSNNRTVIKAGTVLGGKGNNDTEKLTGQVGVGD 1759  
QY 1690 KDGNAKDLSNVWVKTKQKQKSKALLATYNAAGOTNYVTNNPAAEIDRINEQIRFFHVN 1749  
Db 1760 KDGNAKDLSNVWVKTKQKQKSKALLATYNAAGOTNYVTNNPAAEIDRINEQIRFFHVN 1819  
QY 1750 DGNOEPVQVRNGIDSSASCKHSVAIGFOAKADGEAAVAIGRQTOAGNOSIAIGDNAQAT 1809  
Db 1820 DGNOEPVQVRNGIDSSASCKHSVAIGFOAKADGEAAVAIGRQTOAGNOSIAIGDNAQAT 1879  
QY 1810 GDQSIAGTGNVYAGHSGAIGDPSTVKADNSYSGVNNQFTDATOTDFVGVGNNTVTTE 1869  
Db 1880 GDQSIAGTGNVYAGHSGAIGDPSTVKADNSYSGVNNQFTDATOTDFVGVGNNTVTTE 1939  
QY 1870 SNSVALGSNSAISAGTHAGTQAKKSDGTAGTTTAGATGTVKGFAGTAVGAVSVGASGA 1929  
Db 1940 SNSVALGSNSAISAGTHAGTQAKKSDGTAGTTTAGATGTVKGFAGTAVGAVSVGASGA 1999  
QY 1930 ERRIQNVAAGEVSATSDAVNGSOLYKATOSIANATNELDHRHONENKANAGISSAMAM 1989  
Db 2000 ERRIQNVAAGEVSATSDAVNGSOLYKATOSIANATNELDHRHONENKANAGISSAMAM 2059  
QY 1990 ASMPQAYIPGRSMVTGGTATNGOGAVAVGLSKLSDNGQVFKINGSADTQGHVGAAGA 2049  
Db 2060 ASMPQAYIPGRSMVTGGTATNGOGAVAVGLSKLSDNGQVFKINGSADTQGHVGAAGA 2119

QY	2050 GFHF 2053	DB	2120 GFHF 2123
QY	467 LKFTDNSIATKGTTRITKKKICFACCTNDGVDESKPYL---	DB	509 LTIIPAGNG - GTTGTNTISVTKDGIRAGNAITNVASGLRAYDDANFDVLLNSATDLNRH 567
QY	519 -----SGSITVNNNTTGNKQIOVGANGIKFATVANNVANTSATVGTARITEKIGF-AGTN 572	DB	568 VEDAYKGLNLLNEKNANKOPLV-----TDSTAATVGDLR-----KLGWVSTK 610
QY	573 DGVDECAPYLDK - ERLKVGVRVEITDSDGINAGNHKHTGTGTNGTANT-----DAVTI 622	DB	611 NGTKEESNOVKQADEVLFTGAGAAVTYSKSENGKHTT---TVSVAETKADCGLEBKDGDTI 667
QY	623 KOLKDAKPTLNAGDGISINSNGDLVDSSNGIITPTTYNISVTKTTLKASNGTSGNKKFSVS 682	DB	668 KLVKVDNQNT---DNVLTGVNNGTAVTKGGFET-----VKTGATDAD---RGKVTVK 712
QY	683 NAHDNNS---LVYAKOLADLYLNKVNETAADSAALPSFKVQNGDNS---NNAITVGVKDTNGKT 736	DB	713 DATANDADKKVATVKDVATAIN-----SAATFVKTEINTLTSTIDEDNPTDNGKDDALKA 765
QY	737 FNTLKLK-GEN-----GVNIT-----TNRAGTGTTFGIDQSNGLT--TPKLTVG 777	DB	766 GDTLTFKAGNLKVKRDKGNITFDLAKNLEVKTAKVSDTLTIGNTPTGGTATPKVNIT 825
QY	778 SDTNGNRLVIEQVPSADGNSTKNI - IKGLSPTL--PSIASPSGRNIALGNTIEBKDSNA 834	DB	826 STADGLNFAKE---TADASGKNVYLKGIATLTLEPSAGAKSSH---VDLVNDAATKKSNA 879
QY	835 ASIDDLVNLNAGFLNKNNGKDKDFVSTVDYFDIGNATTATVYDEANOTSKVAYDVNVD 894	DB	880 ASIEDVLRAGWNIQGNVNDYVATVTVNFVTDSTGTTTWTVTQ---KADGKGADVKA 937
QY	895 KTIELTGDNCKKOLGVKTKLTETSTNGNATFTSTDDH-----ALVKASDI----- 941	DB	938 KTSVIKDHNGKLFEG - KDLC-----DANGATVSEDGDKGTGLTAKTVIDAVNKS 991
QY	942 -----AGNLNTLAE--EIHHTKGTANTALQTFVTKVVDEN----- 974	DB	992 RVTGEGATAETGATVAVNAGNAETVTSVFNKGNATTA---TVSKDGNINVKVDVN 1047
QY	975 -----DK---ADDTNATVKGDTSGKV-----NTLUKKGKNGL----- 1005	DB	1048 VGDGLKIGDDKKIVADTTTLTV---TGGKVSVPAGANSVNNKKLVNAEGLATALNLS 1103
QY	1006 -DLKTKDKGTVTFTGINTQSLKAGDSTTLN - NGLSLTN-----TASNEQIOVGADGVKF 1058	DB	1104 WTKADKYADGESEGETDQEVKAGDKVTFKAGNKLKQSEKDFYLSLODTLTGLTSITL 1163
QY	1059 AMVNNVWGAGIDGTTTRITRDEIGFTGTNGS---LDKSKPH---LSKDGINAGKKKTTNI 1112	DB	1164 GGTANGNDTG---IVINKDGLTITLANGAAAGTASNGNTISVTKDGISAGNKEITNV 1219
QY	1113 QSG-----EIAKNSHDVAVTGGKIYDLKTELENKISSTAKTAQNSL 1152	DB	1220 KSALKTYKDTQNTADETQDKEFHAUVKNAEVEPVG-----NGATVSAKTDNNGK 1270
QY	1153 H-----EFSVADEQGNFTVSNPYSSYDTSK-----TSDVI 1183	DB	1271 HTVTIDVAEAKVGDLKDPDGKIKLKVDNTDGNLLTVDATKASVAKGEFNAVTTDAT 1330
QY	1271 HTVTIDVAEAKVGDLKDPDGKIKLKVDNTDGNLLTVDATKASVAKGEFNAVTTDAT 1330	DB	1331 TAQGTNANERGVVVKSGNSGATATETDKKVVATVDGVAKAINDAATFVKVENDDSATIDD 1390
QY	1391 SPTDDCANDALKAGDTLTLKAGNKLKVKRDKGNITFPALANDLSVKSATVSDKLSLGTNGN 1450	DB	1451 KVNITSDTKLGNFAKDSKTKTGDGDANIHLNGTASTLDTFLLNSGAT-TNLGGNGITDNEKKR 1509



Db 380 GLKFDSDKKIIVADTALTAVTGK--VAEIAKEDD-----KKLVNAG----- 419  
QY 353 LKLVELANRKITPKGDGNN-----SNSVERGLGNTLTIKGDAQTNALTEANIGVVYTD 406  
Db 420 -DLVTALCNLSWAKABADTDGALEGISKQDEKAGETVTFKAG-----KNLKVQOD 470  
QY 407 GNLKVKLAKELTGLTSVSATNKLTVSNNTNNNAELQSGGLTFPSPITGCTKTDKTVYSIDG 466  
Db 471 GANTYSLQDALTLGLTS-----ITLGTTN-----GGNDAKTVINCKD 508  
QY 467 LKFTNDSNSIATGCTTRITKKIKIFAGTNDGVDESKPVL-----DNEKLKVGNST---LN-- 518  
Db 509 LTIIPAGNG-CTTGNTISVTKDIKAGNKAITNVASCLRAYDDANEDVLNNSATDLNRH 567  
QY 519 -----SGSLAVNTTGNKQIQVANGIKFATVANNVANTSATGARTERIKGF-AGTN 572  
Db 568 VEDAYKGLLLNLNEKNANKOPLV-----TDSATAVTGDLR---KLGWVYSTK 610  
QY 573 DGVDEQAPYLDK--ERLKVGRVELITDGSINAGNHKITGLTNGIANT-----DAVTI 622  
Db 611 NGTEESNQVKQADDEVLTGAGAAVTVSKSENGKHTI---TVSAETKADCGLEKDGDTI 667  
QY 623 KOLDAKPTLNAGDGISINSNNGDLVDSSGNIITPTVYNISVKTTLKLSNGTSGNNKESVS 682  
Db 668 KLVVDNQMT---DNVLTVGNNGTAVTKGGPET-----VKTGATDAD-----RGKVTVK 712  
QY 683 NAHDNNS---LVTAKDLADLYLNKYNETADSALPSFKVQNGDNS---NNAITVGRKDTNGKT 736  
Db 713 DATANDADKKVATVKDVATAIN-----SAATFVKTEIENLTTSIDEDNPTDNGKDDALKA 765  
QY 737 PNTLKLK-GEN-----GVNIT-----TNRATGTVTFGIDQDSNGLT--TPKLTUVG 777  
Db 766 GDTLTFKAGKNLKVGRDKGNTFLDLAKNLEVKTAKVSDTLTIGGNTPTGGTTATPKVNIT 825  
QY 778 SDTNGNRLVIEOPVSDAGNSPKNT-IGLSPTL--BSTASPSGRNIALGNTIEEKDSNA 834  
Db 826 STADGLNFAKE---TADAGSKNNYLKGIATLTPEPSAGAKSSH---VDLNVDATKKSNA 879  
QY 835 ASIDDLVLNAGNLKNNGKDKDFVSTYDVFIDGNATATVTYDEANOTSKVAYDVNVYDE 894  
Db 880 ASIEDVLRAIGNIQQNGNVYDVATYDVFNTDSDTGTGTTVTVQ--KADCKGADVKIGA 937  
QY 895 KTIELTDGNGKKQLGVKTIKLTETSTNGNATFTSTDDH-----ALVRASDI----- 941  
Db 938 KTSVIKOHNGKLFEG-KDLK-----DANNGATVSEDDGKDTGLTGLTAKTVIDAVNKS GW 991  
QY 942 -----AGNLNTLAE--EIHHTTKGTANTALQTFVKKVDEN----- 974  
Db 992 RVTGEGATAETGATAVNAGNAETVTSCTSVPFKNGNATTA-----TVSKDNGNIWYKDVN 1047  
QY 975 -----DK--ADDTNAITVGRDGTSGKY-----NTLKGKNGL----- 1005  
Db 1048 VGDGLKIGDDKKIIVADTTLTV-----TGGKVSVPAGANSVNNKKLVNAEGLATALNLS 1103  
QY 1006 -DIKTDKDTGVTFGINTQSGLKAGDSITLN--NGLSITKN-----TASNEQIQVGADGVKF 1058  
Db 1104 WTAKADKYADGESEGETDQEVKAGDKVTFKAGKNLKVQSKDEFTYSLQDTLTGLTSITL 1163  
QY 1059 AMVNVGVGAGIDGTRITREIGFTGTPNGS---LDKSKPH---LSKDGINAGKKIINI 1112  
Db 1164 GGTANGRDTC---TVINKDGLITLANGAAAGTDAENGNTISVTKDGISAGNKEITNV 1219  
QY 1113 QSG-----EIAKNSHDAVTGGKIYDLKLELENKISSITAKTAQNSL 1152  
Db 1220 KSALKTYKDTQNTADETDQKEFFAAVKNANEVEFVGK-----NGATVSAKTDNNGK 1270  
QY 1153 H-----EFSVADEQGNFTVSNPSSYDTSK-----TSDVI 1183  
Db 1271 HTVTIDVAEKVGDLEKDKTDGKIKLKVNDNTDGNLLTVDATKGAASVAKGEFNAVTTDAT 1330  
QY 1184 TFAENG-----ITTKVNGVVRVIDQTKGLT----- 1211

Db 1331 TAQGTNANERKVVVVGSGNGATATETDKKKVATVCDVAKAINDAATFFVKVENDSDSATIDD 1390  
QY 1212 -----TPKLTVGNNG 1222  
Db 1391 SPTDDGANDALKAGDTLTLKAGKNLKVRRDGNITFALANDLSVKSATVSODKLSLGTNGN 1450  
QY 1223 KGINVSONGON-----TITGLSNTLANVTNDKSGSVTTTBOGNIILKDEDKTR 1269  
Db 1451 KVNITSDPKGLNFRAKDSKTGDODANIHNLGIASTLTDTLNLSGAT--TNLGGNGIITNEKR 1509  
QY 1270 AASIVDLVLSAGFNLOG-----NGEAVDFVSTVDTVNFANGNTTTAKTVYD--DTSKT 1319  
Db 1510 AASVKDLVLNAGNVRGVKVPASANNQVENIDPVAITDVTDFVSGDKDTTSVTVESKDNGKR 1569  
QY 1320 SKVVYDVNVDDTIEVKDKKLGKVTTLTSTG-----TGAN-KFALSNOAT-----G 1365  
Db 1570 TEV-----KIGAKTSVIKDHNGKLTGKELKADANNNGVTVTETDGKDEG 1613  
QY 1366 DALYKASDIVAHLNLTLSGDIQTAKASOANNSAGYVD-----ADGNKVIVYDSTDNKKY 1418  
Db 1614 NGLVTAKAVIDAVNKAAGRVRKTT--GANGONDDFATVASGNTVTFADGNGTTAEVT----- 1667  
QY 1419 QAKNDG--TVDKTKEVAK-----DKLVAQA---QTPDGTLAQMNVKSVINKEQVNDANK 1467  
Db 1668 -KANDGSTIVKYKYNKVADGLKLDGKIVADTIVTLTVADGKVTAPN-----NGDCK 1716  
QY 1468 KOGINEDNAFYKLEKAASDNKRNAAVTVGLDNVAOATPLTFFAGDTGTGTAKKLGELTIT 1527  
Db 1717 K-----FVDSAGLADALNKLSWTA--TAGKEGTGEVDPANSAGQ-----EVKAGDKVTF 1763  
QY 1528 KGGQDITNKLTDNNIGVGVAGTDGFTVKLAKOLITNLSV-----NAG-----GPKIDEKGIS 1578  
Db 1764 KAG-----DNLKIKQSGKDEFTYSLKELKDLTSEFEKDANGGTGSESTKITKDGULT 1814  
QY 1579 FVDANGQA---KANTPVLPSANGLDLGKVSINVGKTF-----DIDAAN 1619  
Db 1815 ITPANGACAAGANTANTISVTKDGISAGNKAVTNVSGLLKFGDGHTLANGTVADFEKHY 1874  
QY 1620 VQOLNEVRNLLGLGNDN---ADGNQVNIAD-----IKKDPNSGS-----SSN 1658  
Db 1875 DNAYKDLTNLDEKAGDNNPTVADNTAATVGLRLGLGWISADKTTGEPNOEYNAOVNRAN 1934  
QY 1659 RTVIKAGT-----VLGG-----KNNDTTEKLTAG---VOVG----- 1687  
Db 1935 EVKFKSGNGIINVSGKTLNGTRVITFELAKEGVYKVSNEFTVKNADGSETNLVKVGDMYYSK 1994  
QY 1688 VDKDGNANGDLSNVWVKTK-----DGSKALLATYNAAGOTNYVTNNPAAEATD 1736  
Db 1995 EDIDPATSKPMTC---KTEKYKVENGVVVSANGSKTEVTLFNKSG---YVTGN--QVAD 2046  
QY 1737 RINEQIRFFHYNDGNQBPVVOGRNGIDSSASGKHSVAIGFOAK-----ADGEA 1785  
Db 2047 AIAKSGFEL-----GLADAAEAFAESAKDKQLSKDKDAETVNAHDVK 2090  
QY 1786 AVAIGRQTOAGNOSTAIGDNAQATGDO-----SIAIGTGNVAGKHSIGAIDPSTVK 1837  
Db 2091 RFANGLNTKV---SAATVESTDANGDKVTTTFVKTDELPLTQIY---NTDANGKNIVKK 2144  
QY 1838 ADNSYSVGNNGQFTDATOTDVFVGNNITVTESNSVALGCSNAISAGTHAGTQAKKSDGT 1897  
Db 2145 ADGKW-----YELNADGTASNKEVTLGNVDANGKK-----VVKVTENG 2182  
QY 1898 AGTTTTAGATGTVKGFAGQAVGAVSVG-----ASGABERRIONVAAGEVSATS 1945  
Db 2183 ADKWTYNADGAADTKGEVSNDKYSTDEKHVVRVLDPNQNSNGKGVINDVANGEISATS 2242  
QY 1946 TDAVNGSOLY---KATQSIANATNEDLHRIHONENKANAGISSAMAMSPQAYIPGRSM 2002  
Db 2243 TDAINGSOLYAVAKGVTLNLAGOVNLEKGVKVRADAGTASALASOLQOATPMGRSM 2302  
QY 2003 VTGTTATINGOGAVAGLSKLSNDNGQWVFKINGSADTOGHVCAAVGAGFHF 2053  
Db 2303 VAIAGSSVQGGQNGLAIGVSRISDNKGVIRLSGTTNSQOKTGVAAGVGTOW 2353





Db 1276 VAEAKVGDLEKDTGKIKLVNDTNGNLLTVDTATKASVAKGEFNAVTTDAATTAQGTN 1335  
Qy 1190 G-----ITTKVNGVVRVGIIDQTKGLT-----1211  
Db 1336 ANERKGVVVGKSGATATETDKKVATVGDVAKAINDAATEFVKVENDDSATIDDSPTDGG 1395  
Qy 1212 -----TPKLVGNNNGKGIIVN 1228  
Db 1396 ANDALKAXDTLTKAGKNLKYKRGKNTTFALANDLSVKSATVSDKLSLGTNGKNKVNITS 1455  
Qy 1229 SONGON-----TITGSLNTLANVTNDKGSVRTTEQGNLIKDEKTRAASIVD 1275  
Db 1456 DTKGLNFAKDSKTGDDANIHLNGIASTLTDILLNSGAT-TNLGGNGIYDNEKKXAAASVKD 1514  
Qy 1276 VLSAGNFIQ-----NGEAVDFVSYDVTNFANGNTTTAKVTYD--DTSKTSKVVYD 1325  
Db 1515 VLNAGMNVVRGKVPASANNQVENIDFVATYDVFVSGDKDTSVTVESKDNCKRTEV-- 1571  
Qy 1326 VNVDDTTIEVDKDKLVKVTTLTSTG-----TGAN-KFALSNOAT-----GDALVKA 1371  
Db 1572 -----KITAGTSVWIKDHNGKLFYCKELKDANNNGVTVTETDGRKDEGNGLVTA 1618  
Qy 1372 SDIVAHNLNTSGDIOTAKGASOANNSAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTK 1431  
Db 1619 KAVIDAVNAKGRVKT--GANGQNDFAIV-ASGTNVTF-----ADNGT---TAE 1664  
Qy 1432 VAKDKLVAQAQTPDGLT-AQNMVKSVINKEQVNDANKKQG---INEDNAFVKGLEKAASD 1487  
Db 1665 VTK-----ANDGSITVKYVK-----VADGLKLDGKIADVTLVTVADGKVTAP 1709  
Qy 1488 N-----KTKNAVTVGDLNNAVAQPLTFAGDTGT-----TAKKLGTELTKGGQT 1532  
Db 1710 NNGXKGFEXDASGLAGCLNKLST--ATAGRECTGEVDPANSAGQEVKAGDKVTFKAG-- 1765  
Qy 1533 DTNKLTDNNIGVAGTDCFTVKLAKDLNLSV-----NAG-----GTXIDKGLSFEVDAN 1583  
Db 1766 -----DNLKIKQSKXKFTVSLKELKDLTSVEFKDANGGTGSESTKITKDGGLTITPAN 1818  
Qy 1584 GOA-----KANTPVLISANGLDIGKGVISNKGKTKDTPDAANVQOLNEVRNLLGLGNDNAD 1638  
Db 1819 GAGAAGANTANTISVTKDGISAGNAKAVTVNVSGLKK-----FGDGHITLAN 1863  
Qy 1639 GNOVNIADIKK 1649  
Db 1864 G---TVADFEK 1871

RESULT 5  
US-08-685-467-4  
; Sequence 4, Application US/08685467  
; Patent No. 6060059  
; GENERAL INFORMATION:  
; APPLICANT: St. Gene III, Joseph W.  
; APPLICANT: Barenkamp, Stephen J.  
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/685,467  
; FILING DATE: 22-JUL-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/409,995  
FILING DATE: 24-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-61053-2/RET/RMS  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1912 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-685-467-4

Query Match 8.6%; Score 890; DB 3; Length 1912;  
Best Local Similarity 23.8%; Pred. No. 3.5e-45;  
Matches 502; Conservative 227; Mismatches 680; Indels 702; Gaps 97;

Qy 1 MNHIYKIPNKGTFMVAEYAKSHSTGGSCATQGVGS-----VRTLSFARIAALA-- 53  
Db 1 MNKIPNVIWNVMTQWVVSELTRTH-----TKRLNRNGDPVLATLLFPATVQANATD 52  
Qy 54 -----VLVIGA----- 59  
Db 53 EDEELDPVVRAPVLSFHSKDEKTEKEVENTENNGIYFDNKGVLKAGAITLKAGDNLKX 112  
Qy 60 -----TLNGSAYAOOI-----TTKTEIG----- 77  
Db 113 KOXTDEXTNASSFTYSLAKDLTDLTSVATEKLSFGANGDKVDITSDANGLKLAKTNGNV 172  
Qy 78 QTNKINNTLKGDALATGEASTAFGSLSKAQSQAIAIGSVKPPDPNNGSNGVSHAKGN- 136  
Db 173 HNLGDLSTLPOAVTNTGVLSSSFPTNDVEKTRA---ATVKDVLNAGNWKAKTAGGNV 229  
Qy 137 ESIAIGGDVLAGSDASIAIGSD--DLVLPKNLDLKNFEHKLHGHETLKKLTOTSDGKI 193  
Db 230 ESDVL---VSAYNVVEFITGDKNTLDVVLTAKENAKTTEVRETPKTSVIKE-----KDGKL 282  
Qy 194 ---KYRTRRAOCHASTAVGAMSQAQCHFSNAPGTATAEAYSLAVGLAAQAATKOSIAV 250  
Db 283 FTGKENNDNTKVTSTNTATDNTDEGNG-----LVTAKAVID-AVNKAGWRVKTT--- 330  
Qy 251 GSNKANAFAATAIGNTVNVNLGRVGLFGSQIILDRDNTDA-----SAYVPLG----- 300  
Db 331 -ANGONGDFATVASGNTVTFESGDGTTAS-----VTKDTNGNGITVKYDAKVGDLGLKFD 384  
Qy 301 --KTLADQYKATROGSDTDIFSIGNNNNNSSIRKILNVGAGSRDTPDAVNVQAQLKLEE 358  
Db 385 DKKIVADTTALTVTGKG--VAEIAKEDD-----KKLVNAG-----DLVTA 423  
Qy 359 LANRKITEFGDGDNN-----SNSVERGLGNTLTIKGDAQTNALTEANIGVVVTDGNGLVK 412  
Db 424 LCNLSWKAKAEADTDGALEGISKDEKAGEVTFKAG-----KNLVKQDQGANFTY 475  
Qy 413 KLAKELTGLTSVATNKITVSTNNNNNAELOSGLTFFSPITGKTGKTIVYSIDGLKFTND 472  
Db 476 SLQDALTLGLTS-----ITLGGTTN-----GNDAKTVINKDGLTITPA 513  
Qy 473 SNSIATKGTTRITKKKIGFAGTNDGVDESKPYL---DNEKLVGNST---LN-----S 519  
Db 514 GNG-GTTGTNTISVTGDKIGAKNKAITNVASGLRAYDDANFDVNNNSATDLNRHVEDAYK 572  
Qy 520 GSLTVNNTTGNKQIOVGANGIKFATVANNVANTSATVGTARITEKIGF-AGTNDGVDEO 578  
Db 573 GLLNLNEKNANKQPLV-----TDSATVGDLR-----KLGWVSTKNGTKEE 615  
Qy 579 APYLDK--ERLKVGRVEITTDGGINAGNHKITGLTNGIANT-----DAVTIKQLKDA 628

```

Db 616 SNQVQADEVLEFGAGAAVTSKSENGKHTI---TVSVAETKADCGLEKDGDTIKLVDN 672
QY 629 KPTLNAGDGLSINSNGDLVDSSGNITPTPIYNSVKTTLKNSNGTSGNNKFSVSNADNN 688
Db 673 QNT---DNVLTGVNNGTAVTKGFET-----VKTGATDAD-----RGKVTVKDATAND 717
QY 689 S---LVTAKLADYLNKVNETAADALPSFKVQNGDNS---NNAITVGKDTNGKTFNTKL 742
Db 718 ADKVVATVQVATIN-----SAATEVKTENLTYSIDEDNPTDNGKODALKAGDTLIF 770
QY 743 K-GEN-----GVNIT-----TNRAATGVTEGIDOSNGLT---TPKLTVGSSTNGN 783
Db 771 KAGNKLKVKRGKGNITPDALKNLEVTAKVSDTLTIGGNTPTGTTATPKVNITSTADGL 830
QY 784 RLVEIQVPSADGNSTKNI-LKGLSPTL--PSIASPSGRNIALGNTTEKDKSNAASIDV 840
Db 831 NFAKE---TADAGSKNYYLKGIATTLTETPSAGAKSH---VDLNVDTAKSNAASIEDV 884
QY 841 LNAGFNLNKNGKDRFYSTYDVFIDGNATTATVYDEANQTSKVAYDVNVDEKTIET 900
Db 885 LRAGWNIQGNVVDYVATYDVTNFTDSTGTTVTQ---KADGKADVKIGAKTSVK 942
QY 901 GDNCKKQKGVKTKLTETSTNGNATFSTDDH---ALVKAADI-----941
Db 943 DHNGKLEFG-KDLK-----DANNGATVSEDDGKDTGTLVTAKTVIDAVNKGSRVTEG 996
QY 942 -----AGNLTLAE--EIHHTTKGTANTALQFTVKKVDEN-----974
Db 997 ATAETGATVAGNAEIVTSGTSVFNFGNATTA---TVSKDNGNINVKYDVNVGDLK 1052
QY 975 ---DK---ADDTNATVVGKDTSGKV-----NTLKLKGNGL-----DIKTD 1010
Db 1053 IGDDKKIVADTTTLTV---TGGKVSVPAGANSVNNKLVNABELATALLNLSWTAAD 1108
QY 1011 KGVITVGTQSLKAGDSTTLN-NGLSIKN-----TASNEQIQVAGDVKFAVYNG 1064
Db 1109 KYADGESEGTQVQKAGDKVTFKAGNKLKVKSEKDTYSLODTLGLTSITLGGTANG 1168
QY 1065 VVAGIDGTTTRITDEIGFTNGS---LQSKPH---LSKDGINAGKKTITNQS--- 1115
Db 1169 RNDTG---TVNKDGLTITLANGAAGTDAANGNTISVVKDGISAGKEITNVKSALK 1224
QY 1116 -----EIAKNSHDVATGGKIYDLKTELENKISSTAKTAQNSLH----- 1153
Db 1225 YKDTQNTADETQDKEFAAVKNAEVEFVK-----NGATVSAKTDNNGKHTVID 1275
QY 1154 -----EFSVADQGNFTVSNPYSSYDTSK-----TSDVITFAGEN 1189
Db 1276 VAEKVGDLKEDTDGKIKLVKVDNFDGNNLLTVDTATKASVAKGEFNAVTTDATTAGTN 1335
QY 1190 G-----ITTKVYKGVVRVIGIDTKGLT-----TPKLTGVNNGKGVIN 1211
Db 1336 ANERKVVVKGSGATATETDKKVVATVDVAKAINDAATFVKVENDDSATIDDSPTDGG 1395
QY 1212 -----TPKLTGVNNGKGVIN-----TPKLTGVNNGKGVIN 1228
Db 1396 ANDALKAXDTLTLKAGNKLKVKROGKNTFPALANDLSVKSATVSDKUSLGTNGKVNITS 1455
QY 1229 SONGON-----TITGLSNTLANVYNDKGSVRTTEQGNILKDEKTRAASIVD 1275
Db 1456 DTKGLNFAKDSKTGDANIHLNGIASTLTDLNLSGAT-TNLGGNGITDNEKKXAAVKD 1514
QY 1276 VLSAGFNLOG-----NGEAVDFVSTYDVNFANGNTTTAKVTYD---DPSKTSKVYVD 1325
Db 1515 VLNAGWNVRGVSPASANNQVENIDVATYDVFVSGDKDTSVTVESKNGKRTVE--- 1571
QY 1326 VNVDDTIEVDKKLGVVTTTLTSTG-----TGAN-KFALSNOAT-----GDALVKA 1371
Db 1572 -----KICATSVIKDHNGKLFVGLKELKDANNQVTVTEFDGDEGNGLVTA 1618
QY 1372 SDIVAHNLNLSGDIOTAKGASOANNSAGYVDADGNKVTYDSTDNKYQAKNDGTVDKTK 1431

```

```

Db 1619 KAVIDAVNKAGNRVKT--GANGQNDDFATV-ASGTNVTF-----ADNGGT---TAE 1664
QY 1432 VAKDKLVAQAQPDGTL-AQMNKVSINKQVNDANKKQG---INEDNAFVKGLEKAASD 1487
Db 1665 VTK-----ANDGSIIVKYNVK-----VADGLKLDGDKIVADTTVLTVADGKVTA 1709
QY 1488 N-----KTNNAVTVGDLNNAVAQTPLTTFAGDTGT-----TAKKLGTELTLTKG 1532
Db 1710 NNGXGKKFADAGSLAGLKNLKSXT--ATAGKEGTGEVDPANSAGQEVKAGDKVTFKAG-- 1765
QY 1533 DTNKLTDNNIGVVGAGTGTGFTVKLAKDLTNLSV-----NAG-----GTKIDEXGISFVDAN 1583
Db 1766 -----DNLKKQSKDFTYSLKELKDLTSVEFKDANGGTGSESTKITKGLTITPAN 1818
QY 1584 GOA-----KANTPVLVSANGLDLGGKVISVNGKGTKPTDAANYOQLNEVNNLLGLGNDNAD 1638
Db 1819 GAGAAGANTANTISVTKDGISAGNKAVTNVVSLKK-----FCDGHTLAN 1863
QY 1639 GNOVNIADIKK 1649
Db 1864 G---TVADFEK 1871

RESULT 6
US-08-409-995-2
; Sequence 2, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Gene III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1098 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-409-995-2

```

```

Query Match 6.7%; Score 690; DB 1; Length 1098;
Best Local Similarity 24.7%; Pred. No. 1.6e-33;
Matches 314; Conservative 151; Mismatches 439; Indels 366; Gaps 48;
QY 953 HTTKGTANTA---LQTFVKKVKKVNDKXADDITNATVTKDGTSGKVT-----996
Db 26 HTKCSATVAVAVLATILLSATVEANNPTVTKLKAYGDANFNFTNNSIADAKEQVQEAY 85

```

QY 997 ---LKLKGNGLDIKTDKGTIVTIGINTQSLKAGSDSTTLNNG--LSIKNTASNEQIQV 1051  
Db 86 KGLNLNLEKNASDKLLVEDNT-----AATVGNLRLKGLWGLSSKNGTNRKESQ- 132  
QY 1052 GADGVKTFAMVNVGVVAGIDGTTTRITRDEIGFTGNSLDKSKPHLSKDGINAGCKKITN 1111  
Db 133 ---QVKA-----DEVLFEG-----KGVQVTS 152  
QY 1112 IOSGEIAKNSHDAVVGKIYDLKTELENKISSATAQNSLH--EFSVADPQGNFTVSN 1169  
Db 153 -----TSENGKHTITFAKDLG----- 170  
QY 1170 PYSSYDTSKSDVITFAG--ENGLITT--KYNKGVVRVIGIDQTKGLTTPKLTGVNNGKGI 1225  
Db 171 ----VKTATVSDTLTIGGAAAGATTPPKVNV-----TSTTDGLKFAKDAAGANGDTTV 220  
QY 1226 VINSQNGNTITGLSNTLANVNDKSGVRTEQNIITKDEPK--TRAASIVDVLISAGFNL 1283  
Db 221 HLN-----GIGSILTDPL--VGSPTHIDG--DQSTHYTRAASIKDVLNAGWNI 266  
QY 1284 Q-----GNGEAVDFSVYDVFANFANGNTTAKVTVDDTSKSVVYDVNVDDTIEV 1335  
Db 267 KGVKAGSTGQSENVDFVHTYDTEFLSADPTTTVTVDSEKNGKRTKVEKIGAKTSVIKE 326  
QY 1336 KDKKLGK-KTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTLSGDIQTAKGASQA 1394  
Db 327 KDKLFTGKANKEFKVKGAN--ATEDADEGKGLVTAKDVIDAVNKTGWRIKTTDANGON 384  
QY 1395 NNSAGYVD-----ADGN--KVITYDSTDNKYQAKNDGTVDKTEKVAKDKLVAQAQTPD 1445  
Db 385 GDFATVAGSTNVTASNGTATVINGTDG--ITVKYDAKVGDLGLDKGXIAA-----D 437  
QY 1446 GTLAQMNVKSVINKEQVNDANKKGINEDNFAVKGLEKAASDNK---TKNAAVTVGDINA 1502  
Db 438 TTALTVN-----DGKANPNPKGVADVASTDEKKLVTAKGLVTA--LNS 479  
QY 1503 VAQTPFLTFAGDTGT-----TAKKLGTELTKGGQDITNKLTDNNIGVYAGTGDGFTVKL 1555  
Db 480 LSWTTTAAEADGGTLDGNASEQVEKAGDKVTFKAG-----NLKVQOEGANFTYSL 530  
QY 1556 AKDLTNLSV-----NAGTPKIDEKGISFVDANGQA--KANTPVLISANGLDLGGKVISN 1607  
Db 531 QDALTLGLTSITLGTGNGKAKEINIKDGLTTPANGAGANNANTISVTKDGLISAGQSQVKN 590  
QY 1608 VKGTKDTDAANVOOLNEVRNLLGLGNDNAGNQVNTADIKKDPNSGSSSRTVIKAGTV 1667  
Db 591 VVSGLKFKFGDANFDPITSSADNLTKQNDAYKGLTNLDEKGTDKQTPVVAADTAATVGD- 649  
QY 1668 LGKGNNDTEKLATGGVQGVYDKGNANGDLSNVVKTQDKSKKALLATYNAAGQTNVY 1727  
Db 650 LRGLGWISADKTTGGSTEHYHQVRNANE-----VKF-KSGN-----GINVSGKT--- 693  
QY 1728 TNNPAAEADIRINEQIRFFHVNDG-----NQPVVYQGRNGIDSSASGKHSVAIG--FOAKA 1781  
Db 694 -----VNGRREITFELAKGEVVKSNFTVKETNGKETS-----VKVGDKYYSKE 738  
QY 1782 D-----GEAAVAIGRQTA-----GNOSIATGDNAQATGDQSIAIGTGNVYAGKHSAGIG 1831  
Db 739 DIDLTGTPKLIKDGNTVAAYKODKGGKVVSVTDNTEATITNK-----GSGYVTGNQVADAI- 794  
QY 1832 DPSTVKADNSYSVGNNO-----FTDATQDFVFGVGNINIVTESNVALGSNAISAGT 1885  
Db 795 -----AKSGPELGLADEADAKRAFDDKTKALSAGTTEIVNAHDKVRFANGLNTKVSAA 848  
QY 1886 HAGTQAK-----KSDGTAGTT 1901  
Db 849 VESTDANGKVTITTFVKTDVPLPTQIYNTDANGKKITKVVYKDGQTKWYELNADGTADMT 908  
QY 1902 TT-----AGATGTVKGFAGQATVAVSVG-----A 1926  
Db 909 KEVTILGNVDSGKVKVNDGDKWYHAKADGTADTKGEVSNDKYSTDEKHVVSLDPNDQS 968  
QY 1927 SGAERRQNVAAGEVSATSTDAVNGSGLY---KATQSIANATNELDRIHQENKANAGI. 1983

Db 969 KGKGVVINDVANGDISATSTDAINGSQLYAVAKGVNTNLAGOVNLEGKVNKVGKRADAGT 1028  
QY 1984 SSAMAMASMPQAYIPGRSMVTGGIATHNGOGAVAVGLSKLSDNGOWFKINGSADTQGHV 2043  
Db 1029 ASALAASOLPOATWPKSKVAIAGSSYOGNGLAIGVSRISDNGKVIIRLSGTTNSQSKT 1088  
QY 2044 GAAVAGPHF 2053  
Db 1089 GVAAGVGYQW 1098  
RESULT 7  
US-08-685-467-2  
; Sequence 2, Application US/08685467  
; Patent No. 6060059  
; GENERAL INFORMATION:  
; APPLICANT: St. Gene III, Joseph W.  
; APPLICANT: Barenkamp, Stephen J.  
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/685,467  
; FILING DATE: 22-JUL-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/409,995  
; FILING DATE: 24-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M. 38,304  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-61053-2/RPT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1098 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-08-685-467-2  
Query Match 6.7%; Score 690; DB 3; Length 1098;  
Best Local Similarity 24.7%; Pred. No. 1.6e-33;  
Matches 314; Conservative 151; Mismatches 439; Indels 366; Gaps 48;  
QY 953 HTTKGTANTA---LQTFVTKVVDENDKADDTNATIVGKDGTSGKVNT----- 996  
Db 26 HTKASATVAVAVLATLLSATVEANNTPVTKLKAYGDANFNFTNNSIADAEKQVQRAY 85  
QY 997 ---LKLKGNGLDKTDKDGTVTFGTGINTQSLKAGSDSTTLNNG--LSIKNTASNEQIQV 1051  
Db 86 KGLNLNLEKNASDKLLVEDNT-----AATVGNLRLKGLWGLSSKNGTNRKESQ- 132  
QY 1052 GADGVKTFAMVNVGVVAGIDGTTTRITRDEIGFTGNSLDKSKPHLSKDGINAGCKKITN 1111  
Db 133 ---QVKA-----DEVLFEG-----KGVQVTS 152

us-09-361-619-9.ra1

Thu Sep 13 14:19:21 2001

```
QY 1112 IOSGEIAKNSHDAVTGGKIYDLKTELENKISSTAKTAQNSLH--EFSVADQGNFTVSN 1169
Db 153 -----TSENGKHTITFALAKDLG----- 170
QY 1170 PYSSYDTSKTSVITFAG--ENGIT--KVNGVVRVGDIDQKGLTTPKLTGVNNGKGI 1225
Db 171 ----VKTATVSDTLTGGGAAGATTTPKVN-----TSTDGLKFAKDAAGANGDTTV 220
QY 1226 VINSQNGONTITGLSNFLANVTNDKGSVTTBOGNIIDEDK--TRAASIVDLSAGFNL 1283
Db 221 HLN-----GIGSLTDTL--VGSPTHIDGG--DQSTHYTRAASIKDVLNAGWNI 266
QY 1284 Q-----GNGEAVDFVSYDVFNFANGNTTAKVYDDTSKTSKVVDVNVDDTTIEV 1335
Db 267 KGVKAGSTTGOSENVDFVHYDVFELSDATETTTVTVDSKENGKRTVEKIGAKTSVIRE 326
QY 1336 KDKKLGV-KTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSGDIQTAKGASQA 1394
Db 327 KDKLFTGKANKETNKVDGAN--ATEDADEGKGLVTAKDVIDAVNKTGWRIKTTDANGON 384
QY 1395 NNSAGYVD-----ADGN---KVIYDSTDNKYQAKNDGTVDKTKVAKDKLVAQAOTPD 1445
Db 385 GDFATVASGNTVTFASGNGTTATVTNGTDG---ITVKYDAKVGDLKLDGDKIAA-----D 437
QY 1446 GTLAQMNVKSVINKBQVANDANKKQGINEDNAFVKGLEKASDNK---TKNAVTVGDNA 1502
Db 438 TTALTIVN-----DGKNANPNKGVADVASTDEKKLVTAAGLVTAA--LNS 479
QY 1503 VAQTPLETFAGDTGT-----TAKKLGTLTIGKQDNTINKLTDNNIGVVAGDTGGTVKL 1555
Db 480 LSWTTTAAEADGTLGDNASQEVKAGDKVTFKAGK-----NLKVKQEGANFTYSL 530
QY 1556 AKDLNLNSV-----NAGGKIDEKGLSFVDANGQA--KANTPVLNANGLDLGGKVISN 1607
Db 531 QDALTLGLTSITLGTGNGAKTEINKDGLTITPANGAGANNANTISVTKDGLSAGGQSVKN 590
QY 1608 VGGKGTQDAAVQOLNEVRNLLGNDNADGNNOVNADIKDPNCSGSSSRNRTVIKAGTV 1667
Db 591 VVSGLKFGDANFDPLTSSANDLTKQNDAYKGLTNLDEKDTKQTPVADNTAATVGD- 649
QY 1668 LGKGNDNDEKATLGATGVQGVQDKNANGDLSNVWVKQKDGSKKALLATYNAAGQTNV 1727
Db 650 LRLGLWVISADKTTGGSTHEYHQVRNNE-----VKF-KSGN-----GTVSGKT--- 693
QY 1728 TNPAEALDRINOGIRFEHVNDG-----NOEPVQGRINGIDSSAGKHSVAIG--FOKA 1781
Db 694 -----VNGRREITFELAKGEVVKVSNFETVRETNGKETS-----VKVGDKYYSKE 738
QY 1782 D-----GEAAVAITGROTA-----GNOSTAIGDNAQATGDQSIAGTGNVAGKHSAGIG 1831
Db 739 DIDLTGPKLKDGTVAARYQDKGVVSVTDNTEATITNK---GSGVTVGNQVADAI- 794
QY 1832 DPSTVKADNSYSVGNNO-----FTDATQTDVFGVGNNTVTESNVALGSNSAISAGT 1885
Db 795 -----AKSGFELGLADEADAKRAFDDKTKALSAGTTEIVNAHDKVRPANGLNTKVSAA 848
QY 1886 HAGTQAK-----KSDGTAGTT 1901
Db 849 VESTDANGDKVTTTFVKTDVDELPIQTYNTDANGKTKTKVVKDQGTQKWTLELNADGTADMT 908
QY 1902 TT-----AGATGTGVKGFAGQTAAGVAVSG-----A 1926
Db 909 KEVTLGNVDSGKVVVDNDGKWTYHAKADGTADTKTKEVSNDKVSTDEKHVVSLDPNDQS 968
QY 1927 SGAERRIONAAGEVSATSTDAVNGSOLY---KATQSIANATNELDHRHONENKANAGI 1983
Db 969 KCGKGVVDNVANGDISATSTDAINGSQLYAVAKGVTVNLAGOVNLEKGVNKKVCKRADAGT 1028
QY 1984 SSAMAMASMPQAYIPGRSMWTTGGTATINGQCAVAVGLSKLSDNGQWFKYKINGSADTQGHV 2043
Db 1029 ASALAASQLPQATIMPCKSNVAIAGSSYGQGLAIGVSRISDNGKVIIRLSGTTNSQOKT 1088
```

```
QY 2044 GAAVGAGPHF 2053
Db 1089 GVAAGVGYOW 1098
RESULT 8
US-09-377-155-32
; Sequence 32, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, B. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 1098
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-377-155-32
Query Match 6.7%; Score 690; DB 4; Length 1098;
Best Local Similarity 24.7%; Pred. No. 1.6e-33;
Matches 314; Conservative 151; Mismatches 439; Indels 366; Gaps 48;
QY 953 HTTKGTANTA---LOTFFVKVDENDKADDTNATITVGKDTSGKVT----- 996
Db 26 HTKCSATVAVAVLALTLASATVEANNTPVINKLKAYCDANFNFTNSIADAEKQVOEAY 85
QY 997 ---LKLKNGKGLDIDKTDGVTGINTQSGKAGDSTTLNNG--LSIKKTASNEQIOV 1051
Db 86 KGLLNLEKNASDKLLVEDNT-----AATVGNLRKLGWLVSSRNGTRKESQ- 132
QY 1052 GADGVKFAVNVGVVAGIDGTTRITRDEIGFTGTNGSLDKSKPHLSKDGINAGGKKTIN 1111
Db 133 ---QVXHA-----DEVLEFEG-----KGGVQVTS 152
QY 1112 IOSGEIAKNSHDAVTGGKIYDLKTELENKISSTAKTAQNSLH--EFSVADQGNFTVSN 1169
Db 153 -----TSENGKHTITFALAKDLG----- 170
QY 1170 PYSSYDTSKTSVITFAG--ENGIT--KVNGVVRVGDIDQKGLTTPKLTGVNNGKGI 1225
Db 171 ----VKTATVSDTLTGGGAAGATTTPKVN-----TSTDGLKFAKDAAGANGDTTV 220
QY 1226 VINSQNGONTITGLSNFLANVTNDKGSVTTBOGNIIDEDK--TRAASIVDLSAGFNL 1283
Db 221 HLN-----GIGSLTDTL--VGSPTHIDGG--DQSTHYTRAASIKDVLNAGWNI 266
QY 1284 Q-----GNGEAVDFVSYDVFNFANGNTTAKVYDDTSKTSKVVDVNVDDTTIEV 1335
Db 267 KGVKAGSTTGOSENVDFVHYDVFELSDATETTTVTVDSKENGKRTVEKIGAKTSVIRE 326
QY 1336 KDKKLGV-KTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSGDIQTAKGASQA 1394
Db 327 KDKLFTGKANKETNKVDGAN--ATEDADEGKGLVTAKDVIDAVNKTGWRIKTTDANGON 384
QY 1395 NNSAGYVD-----ADGN---KVIYDSTDNKYQAKNDGTVDKTKVAKDKLVAQAOTPD 1445
Db 385 GDFATVASGNTVTFASGNGTTATVTNGTDG---ITVKYDAKVGDLKLDGDKIAA-----D 437
QY 1446 GTLAQMNVKSVINKBQVANDANKKQGINEDNAFVKGLEKASDNK---TKNAVTVGDNA 1502
Db 438 TTALTIVN-----DGKNANPNKGVADVASTDEKKLVTAAGLVTAA--LNS 479
```



FILING DATE: 24-MAR-1995  
PRIOR APPLICATION DATA: PCT/US96/4031  
APPLICATION NUMBER: 22-MAR-1996  
FILING DATE: 22-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Vance, Dolly A.  
REGISTRATION NUMBER: 39,054  
REFERENCE/DOCKET NUMBER: A-61053-1/RET/RMS/DAV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 679 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-913-942-15

Query Match 5.6%; Score 583.5; DB 4; Length 679;  
Best Local Similarity 23.9%; Pred. No. 2.1e-27;  
Matches 226; Conservative 115; Mismatches 223; Indels 383; Gaps 34;

Qy 1135 TELEN-KISSTAKTAQNSLHEFSVAD-----EQNNFTVS-----NPYSSYDTSKTSDDVI 1183  
Db 88 TEVINLNTDSSGNVGSSTITFKAGDNLKTKQSGNDEFTYSLKKELNLTSVETKLS--- 144  
Qy 1184 TFAGENGITTKVNGVVRVGDIDTKGLTTPKLT-VGNNGKGVINSQNGONTITGLSNT 1242  
Db 145 --FGANG-----NK-----VDITSDANGLAKLTGNGG-----QNSNVHLNGIAST 184  
Qy 1243 LANVT--NDKGSVRTTEOGNIKDEKTPRAASIVDVLSAGENLOGNEAVDFVSTYDTVN 1300  
Db 185 LDTLGGTTGTHVDN-----IDAVNYHRAASVQDVLNNGNIOGNNVDFVRYTIDVD 239  
Qy 1301 FANGNTTTAKVTYDDTSKTSKVYDVNVDDTTIEVKDKKLGKVTKTTLTSTGTGANKFALS 1360  
Db 240 FVNGANANVSVDATTAHKT-----TVRVDVTGLPVQ----- 271  
Qy 1361 NOATGDALYKASDIVAHLNLTSLGDIQTAKAGSQANNSAGYVDADGNKYIDSTONKYYQA 1420  
Db 272 -----KTK----- 311  
Qy 1421 KNDGTYDKTKEVAKDKLVAQAOTPDGTLAQMNVKSVINKEOVNDANKOGINEDNAFVKG 1480  
Db 291 KDDGSADMMQKV-----ENGELA----- 308  
Qy 1481 LEKAASDNKTKNAAVTVDGLNVAQAOTPLTFAGDTGTAKKLGSETLTIKGGQDTNKLTDN 1540  
Db 309 -----KTK----- 311  
Qy 1541 NIGVAGTDGFTVKLAKDLTNLSNVNAGTKIDEKGISFVDANGQAKANTPVLNGLDL 1600  
Db 312 -----VKL-----VSASGT-----NPV----- 323  
Qy 1601 GGVISNVGKGTGTDDAANVQOLNEVRNLLGLNDNADGNQVNIADIKDPNSGSSSNT 1660  
Db 324 ---KISNVADGTDGTDVAFKQLKALQD-----KQVTLSTSNAYANGGTDND-- 367  
Qy 1661 VIKAGTVLGGKGNNDTEKLGTVGVQGVVD-KDGNANGDLSNVVWVKTKDQGSKKALLATYN 1719  
Db 368 -----GGK-----ATQTLNGLNFKPKSSDGLLKI----- 394  
Qy 1720 AAGTNTVYNNPABEADIRENEQGRFFHVDNGNOEPVVGVRNGIDSSASGKSHSVATGFQA 1779  
Db 395 ATGDT--VTFTPK----- 412  
Qy 1780 KADGEAAVATGRTQAGNOSTAIGDNAQTQDSIAIGTGNVVGKHSKA-----IGD 1832  
Db 413 -DDGKASISKGANTEG--LVEASELVELSNLKLGWKVGVEKVGSGELDTGTSKETLVKSGD 469

Qy 1556 AKDLTNLSV-----NAGGTXIDEKISFVDANGQA--KANTPVLNGLDLGGKVISN 1607  
Db 531 QDALTGTLSTLGTGNGAKTEINKDGLTITPANGAGANNANTISVTKDGISAGQSQSVKN 590  
Qy 1608 VGGTGTDTDAANVQOLNEVRNLLGLNDNADGNQVNIADIKDPNSGSSSNTVIKAGTV 1667  
Db 591 VVSLGKKFGDANFDPLTSSADNLTQKNDAYKGLTNLDEKGTDKQTPPVADNATAATVGD- 649  
Qy 1668 LGGKGNNDTEKLGTVGVQGVVDKGNANGDLSNVVWVKTKDQGSKKALLATYNAAQTNVV 1727  
Db 650 LRLGLWISADKTTGGSTEHDOVRANE-----VKF-KSN-----GINVSGKT--- 693  
Qy 1728 TNPAEADIRENEQGRFFHVDNG-----NOEPVVGVRNGIDSSASGKHSVAIG--FQAKA 1781  
Db 694 -----VNGRREITPELAKGEVVKVKSNEFTVVKETNGKETS-----VKVGDKYYSKE 738  
Qy 1782 D-----GEAAVAIGRTQA-----GNQSTAGDNAQTQDSIAIGTGNVVGKHSKAIG 1831  
Db 739 DIDLTGQPKLKDNTVAARYKQDGKGVVSVTDTEATITNK-----GSGYVTGNQVADAI- 794  
Qy 1832 DPSTVKAADNSVSGNNNO-----FTDATQTDVFGVGNNTITVTESNSVALGNSAISACT 1885  
Db 795 -----AKSGFELGLADEADAKRAFDKTKALSAGITTEIVNAHDKVRFANGLNTKVSAA 848  
Qy 1886 HAGTQAK-----KSDGTAGTT 1901  
Db 849 VESTDANGDKVTTTFVKTVDVELPLTQIYNTDANGKKTKTVKVDGQTKVYELNADGTADMT 908  
Qy 1902 TT-----AGATGTVKGFAGQTAGVAVSVG-----A 1926  
Db 909 KEVLGNVDSGKKVVDNDGKWTYHAKADGTADTKTGEVNSVDTEKHVWSLDPNDQS 968  
Qy 1927 SCAERRIONAAGEVSATSTDAVNCQSLY----KATQSIANATNELDHRHIOHENKANAGI 1983  
Db 969 KKGKGVVDVANGDISATSTDAINGSQLYAVAKGVTNLAGOVNLEGGKVNKVKRADAGT 1028  
Qy 1984 SSAMAKASMPQAYIPGRSMVTTGGTATHNGQAVAGLSKLSKDNGQWFKINGSADTQGHV 2043  
Db 1029 ASALAASQLPQATMPEKSWAAGSSYQOGNGLAIGVSRISDNGCKVILRLSGTTNSOKT 1088  
Qy 2044 GAAVGAGHF 2053  
Db 1089 GVAAGVGYQW 1098

RESULT 10  
US-08-913-942-15  
Sequence 15, Application US/08913942  
Patent No. 6200578  
GENERAL INFORMATION:  
APPLICANT: St. Geme, Joseph  
APPLICANT: Barenkamp, Stephen J.  
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,942  
FILING DATE: 29-DEC-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/409,995







Db 443 NVTIAPSASRVELGADRNH--SAEVIKVTLLKNNSTLTLTNTT-----ISNLLKS 493  
QY 543 ATVANNVANTSATVGTARITEEKIGFAGTNDGVDEQAPYLDKERLKVGRVEITDSG---599  
Db 494 AHVVNITARRKLTWNSSISIERGSHLLHSEGGOGGVQIDK-----DITSEGGNLT 545  
QY 600 INAGN-----HKITGLTNGICANTDAVTIKQ-----LKD-----AKPTLNAG-----635  
Db 546 IYSGGWVDVHKNTITLGSGLN---ITTKEGDIAFEDKSGRNNLTITAGGTITSNGSNGFR 602  
QY 636 -DGISINSNGDL--VDS-----SGNLT-----TPTNISVKTTKL-----669  
Db 603 FNNVSLNSLGGKLSFTDSREDRGRRTKGNISNKFDPGTLNLSGTVDISNKAAPKVSFWYRDK 662  
QY 670 -----SNGTSGNNKFSVS-----NAHDNNSLVTAOKLADLYLNKVNETAADSPSKF 715  
Db 663 GRTYWNVTTLNVTSG-SKFNLSIDSTGSGTGPSIRNAE-----LNGI--TFNKA--TFN 712  
QY 716 VONGDNNNAITVG-----KDTNCKTFN-TLKLKGENGVNITTHRATCTVFGIDQSNGL 769  
Db 713 IAQGSTANFISIKASIMPFKSNANYALFENEDISVSGGVSFNKLNAS-----SSNI 762  
QY 770 TTPKLTGSDTNGRLVIEQVPSADGNSTKNI-IKGLSPTLPSIASPSGRNIALGN-TIE 827  
Db 763 QTPGVIIKS-----ONFNVSGGSTLNKAEGETAFASIENDLNUNATGNTITIR 812  
QY 828 EKDSNAASIDVDNLNAGFNKNGKDRDFVSTVDYDFIDGNATTATVYDEANQTSKVA 887  
Db 813 QVE-----GTDSDRVNKGVAACKN-----ITFKGN-----ITGSKATTEIK 850  
QY 888 YDWNVDEKT-IELTGDN---GKOLGVKTKLKTETSTNGNATFTSTDDHALVKASDIAG 943  
Db 851 GNVITNKNTNATLRGANFAENKSPNLT---AGNVINNGNLT-----AG 891  
QY 944 NLNPLAEIHTTKTANTALQTFTVKVVDENDKADDTNALTGVKDGTSKGVNTLLKAGN 1003  
Db 892 SIINIAGNLVSKCANLOAITNYIFNVAGSFDNNGASN-LSIARGGAKFK-----DIINNTS 946  
QY 1004 GLDITDKDGTVTFTGINTQSLGKAGDSTTLNNGLSIKNTASBNQIQVAGDGVKFAVMNN 1063  
Db 947 SLNITNSDTPYRPIKGNISNKSOGD-----LNIIDKKSDAEIOIGN-----989  
QY 1064 CVVCGADTGTTRTRDRDEIGFTGNGSLDKSKPHLSKOGINAGGKIYNIOGSETAKNSHD 1123  
Db 990 -----ISQKEGNLTISDDKYNITNQITIKRAGVEGSDSSEAEANLTIQTKE 1037  
QY 1124 AVTGGKIYDLKLENIKISSAKTAQNSLHEFSVADQGNFTVSNPYSSVDTSKTSDVI 1183  
Db 1038 LKLAG-----DLNISGFNAEITAKNGS-----DLTIGNASGGNADAK--VTFDKVKDSKIS 1088  
QY 1184 TTAGEN---GITTIVKNGVVRVIGIDQTKGLT--TPKLTIVGNN-----N 1221  
Db 1089 T-DGHNVTLNSEVKTSGSSNAGNDNSTGLTISAKDVTNNVNTSHKTNISAAAGNVTT 1147  
QY 1222 KGIVINSQONQNTITGLSNTL-ANVTNDKGSVTTTGGNIILKDEK--TRAASIVDVL 1278  
Db 1148 KEGTINATTGSEVTAQNGTIKGNITISQNVTVTATE--NLVTTENAVINATSGTVNIST 1205  
QY 1279 AGFNLOGNGEAVDVSVDYTVNF-ANGNT-TTAKVTYDDTSTKTSKVVVDVNVDDTTIEVK 1336  
Db 1206 KTGDIKGGIE-----STSGNVNITASNTLKVSNITQDVTVTADA-----1246  
QY 1337 DKKLGVKTTTLTSTGTGANKFALSNOATGDALV--KASDVIHAHLNLTLSGDIQ-TAKGASQ 1393  
Db 1247 ---GALTTTAGST-----ISATTGNANITTKTGDINGKVESSSGTVLVTATGTL 1293  
QY 1394 ANNSAGVYDADGNKVIVYDSTDNKYIQAKNDGTVDKTEKAVAKDLVAQAQPPDGTFLAOMNV 1453  
Db 1294 A---VGNISGNTVITADS-----GKLTSTV-----GSTINGT-----1323  
QY 1454 KSVINKQVNDANKKOGINEDNFAFKLEKAASDNKTKNAAVTVGDLNVAQAQTLTTFAGD 1513

Db 1324 NSVITSSQSGD-----IEGHIISGN-TVNVVTASTGD-----ITIGNS 1358  
QY 1514 TGTAKKLGETLTIKGGQDTNKLTDNNNIGVAGTDFGVKLAKDLTLNLSNVNAGGTAKID 1573  
Db 1359 AKVEAKNGAATLTAESG-----KLTOTGSSITSSNGQTTLTAKDSSIAGNINAANVTLN 1413  
QY 1574 EKGISFVDANGCOAKANTPVLSSANGLDL-----GGKVISNV-----GKG---TKDPTDAAN 1619  
Db 1414 TTGLTTLTGGSKINATSGTLTINAKDAKLDGAASGDRTVVATNATNAGSGNVTAKTSSVN 1473  
QY 1620 VO-OLNEVRNLLGLGNDNA-----DGNQVNIADIKKDPNSGSSSNRTVIKAGTVL---G 1669  
Db 1474 ITGDLNII-NGLNIISENGRNTVRLRCKEIDVKYIQ-----PGVASVEEVIKRVLEKVK 1528  
QY 1670 KGKNDDEKATGQVQ-VGVKDKGNANGDLNWNVYKTKDGSKKALLATYNAAGQNTVYT 1728  
Db 1529 DLSDEERETLAKLGSVAVRFEVPNNA-----ITVNTQNEFTTKPSSQVITISEGRACFSS 1582  
QY 1729 NPAAEADIRINEQG 1742  
Db 1583 NGARVCTNVADDG 1596

## RESULT 13

US-08-728-470-10  
; Sequence 10, Application US/08728470  
; Patent No. 5928651  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728,470  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstreser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-633  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1529 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-728-470-10

Thu Sep 13 14:19:21 2001

```
Query Match      4.98; Score 512; DB 2; Length 1529;
Best Local Similarity 21.86; Pred. No. 1.2e-22;
Matches 386; Conservative 229; Mismatches 674; Indels 478; Gaps 91;

QY 143 GDLAEGDASIAIGSDDL-----YLPKRLDLKNEPHK-----LIIGHE 180
DB 72 GQVFLINPGIITIGKDAIINTNGFTASTLDSINENIKARNFTLEOTKDALAEIVNHG-- 129
QY 181 ILKKTQSTWDGKIY--RRTRAQGHASTAVGAMSVAQGHFSNAGFYATATAAASLAVGL 238
DB 130 ---LITVGRDGSVNLGGKVKKEGVISVNGSGISLLAQ-----KIFISD 171
QY 239 AAOATQSSIAVGSNAKANAFAATAIGGN-----TVVNLGRGVALGFGSOLDRDNNTD 292
DB 172 IINPTIYSIAAPENAEINLGDIFAKGGINVRAATIRNKGK---LSADSVSKDKSGNIV 228
QY 293 ASAYVPLGKTLADQYKATQGSTDFISGN--SNNNSSIRKKLIINVGAGSRDQDAVNV 350
DB 229 LSA-----REGEA-----EIGGVISAQNOQAQKGLMITG-----DKVTL 263
QY 351 AQLKVLLELANRKITPKGDGDNNSVERGLGNTLTIKGDAQTNALTEANIGVVTGDGML 410
DB 264 KTGAVID-----LSKEGETYLGDEREGEK-----NGIOLAKKTTLEKGSTI 307
QY 411 KVKLAKELTGLTSVATNKITVSTNNNAEL--OSGLTFSPITG---TKTDKTVYSIDG 466
DB 308 NVS--GREKGGRAIVMGDIALIDGINAQSDDIAKTGG--FVETSGHDLSIGDDVI--VDA 362
QY 467 LKFTDNSNATKGT---RITKKKIGFAGTNDGVE-----SKPYLDNEKLKVGNS 516
DB 363 KEWLLDPDDVSIEFTLSSGRNNTGENQY--TGDTKESPKGNSISKPTLTNSTLE---QI 418
QY 517 LNSGSLVNNITGKQIQVANGKIEKATVANNVANTSATVGTARIETEEKIGPAGTNDGVD 576
DB 419 LRGSYV--NITANNRIYVNS-----SINLSNGSLTLHTKR-----DGVR 456
QY 577 EQAPYLOKELKVGREVTITDGSINAGNHKITGLTNGIANTDAVTIKOLKDAKPTLNAGD 636
DB 457 INGDITSNEN---GNLTAKAGSWDV--HKNITLGTGFLN-----IVAGD 496
QY 637 GISINSNGD-----LVDSSGNIPTIYNSIVKTKLNSNGTSGNNKFSVNAHDNN 688
DB 497 SVAF--EREGDKARNATQAQITVKNKDKOFRENVISINGTGKGLFI---ANQNN 552
QY 689 -----SLVTKADLADYLN-----KVNETAADSALPSEK-----VQNGD 720
DB 553 FTHKPDGELNIGSIVTINQTTKKDV--KWNASKDSYWNVSSLTNTYQKTFIKFVDSGS 611
QY 721 NNNNAITVGKDTNGKTFNTLKLKGENGVNITNRTATGTVTFGIDQSGNLTTPKLTGSDT 780
DB 612 NSQDLRSSRRSPAGVHEN--GIGGKTFNIGAN---AKALFKL-KPNAATDPKKELPITF 665
QY 781 NGNRLVIEQVPSADGNSNFKNIKGLSPLPSTA---SPSGRNIALG---NTIEEKDKSNA 834
DB 666 NAN-----ITATDSSSVFMDIHANLTSRAGINDMSINITGGDLDFSTSHRNSNA 718
QY 835 ASI--DDVLNA---GFNLKN-----NGKDKDFVSTYDTPFDICNATTATVYDEANQT 883
DB 719 FELKDLTINATGNSFLSKQDKDSFYNEVSKHAINSSHNLITLGN-----VTLGGENS 773
QY 884 SKVADVNVDEK--TELIGDNGKKQLGVK--TIKLTETSTNGNATFTFDDHALVKASD 940
DB 774 SSGTGNINITKANVTLQADTSNNTGLKPKRTLTLCNISVEGNLSLTGAN-----AN 825
QY 941 IAGNLATLAEIHTTKGTANTALQ---TFTVKKVDENDKADDTNATITVGKDGTSG---KV 994
DB 826 IVGNL--SIAED--STFKGEASDNLNITGTFT-----NNGTANINIKQ 864
QY 995 NTLKLGK-----KNGLDIKTDKGTVTFFINTOSGLKAGDSTTLNNGNLSIKNTASNEQIQ 1050
DB 865 GYVKLOGDINNKGGLNITTNASGTOKTIINGNITNEKGD-----LNKNIKADAELQ 916
```

## RESULT 14

US-08-719-641-10

Sequence 10, Application US/08719641

Patent No. 6218141

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESS: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-719-641-10

```

Query Match	4.98;	Score 512;	DB 4;	Length 1529;
Best Local Similarity	21.8%;	Pred. No. 1.2e-22;		
Matches 386;	Conservative 229;	Mismatches 674;	Indels 478;	Gaps 91;

Qy	143	GDVLAEGDASIAIGSDDL-----YLPKNLDLKNFHK-----LIHGE 180
Db	72	GOVELPNEGITICKDALINTNGFASTLDSINENIKARNETLEQTKDLAEIVNHG-- 129
Qy	181	ILKKIQISTGCKIKY--RRTRAQGHASTAVGAMSYAQGHFSNAPGTATYATAEAAYSLAVGL 238
Db	130	--LITVGKGSVNLIGGKVKNEGIVSNGGSIILLAG-----KITISD 171
Qy	239	AAQATKOSIAVGSNAKANAPATAIGN-----TVVNLGRVALFGSQILDRDNFTD 292
Db	172	IINFITYIAAPENALNGDIFAKGGINVRAATIRNKCK--LSADSYSKDKSGNIV 228
Qy	293	ASAVVPLGKTLADQYKATRGQSDTDIFSIGN--SNNNNSSIRKKIINVAGSRPTDAVNV 350
Db	229	LSA-----KEGEA---EIGGIVSAQNOAQKGGKLMITG-----DKVTL 263
Qy	351	AQLKLVBELANRKITYTFKGDGDNNSVERGLGNTLTIKGDAQTNALTEANIGVVTDGNGL 410
Db	264	KTGAVID-----LSGREGGETYLGGERGEGK-----NGIQLAKKTLEKGSTI 307
Qy	411	KVKLAKELTGLTYSVATNKITVSTNNNNAL-OSGGLTFSPITG---TKTDTKVYISD 466
Db	308	NVS-GKEGGGRAIVMGDIALIDGAINAGSDIAKTGG--FVETSGHDSIGDDVIT--VDA 362
Qy	467	LKFNDNSNIATKGT--RITKKIKIFAGTNDGDE-----SKPYLDNEKLKAVGNST 516
Db	363	KEWLDPDDVSIEITLSCRNNTGENQYT-TGDTKESPKGNSIKPPLTNTSLTLE---QI 418
Qy	517	LNSGSLTVNNTGKNQIQVANGIKFATVANNVANTSATVGTARITEEKIGFAGNNGVD 576
Db	419	LRRGSYV--NITANNRIYVNS-----SINLSNGSLTLTKR-----DGVK 456
Qy	577	EQAPYLDKERLKVGRVEITDSDGINAGNHKITGLTINGTANTDAVTIKOLKDAKPTLNAGD 636
Db	457	INGDITSNEN---GNLTIKAGSWDV--HKNITLTGTGLN-----IVAGD 496
Qy	637	GISINSNNGD-----LVDSSGNIPTPTVYISVKTTLKNSNGTSGNNKFSVSNAHDNN 688

1660 TVIKACTVL---GGKGNNDTEKATGGVQ-VGVKDKNGANGDLSNVVKTQKSGKALL 1715  
1447 EVIEARVLEKVKDLSDEBRETFLAKGSAVRFPENNA-----ITVNTONEFTTRPSS 1500  
1716 ATVNAAGQTVYVNTNPAEAIDRINBQ 1742  
1501 QVTISEGKACFPSSGNGARVCTNVADDG 1527

RESULT 15  
US-08-038-682-4  
Sequence 4, Application US/08038682  
Patent No. 5549897  
GENERAL INFORMATION:  
APPLICANT: BARENKAMP, STEPHEN J  
APPLICANT: ST. GEME III, JOSEPH W  
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattiare, Ltd  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/038,682  
FILING DATE: 16-MAR-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: BERKSTRESSER, JERRY W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-293  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1477 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-038-682-4

Query Match 4.5%; Score 469; DB 1; Length 1477;  
Best Local Similarity 21.3%; Pred. No. 4, 4e-20;  
Matches 399; Conservative 244; Mismatches 665; Indels 568; Gaps 91;

QY 1 MNHUYKVFENKATGTETMAVAEYAK--SHSTGGSCATQGVSGVRLTSFARIALAVLVIG 58  
DB 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKGEKSEKPARM-KVRHLAKPLSAM-LLSLG 58  
QY 59 ATLNGSAYAQIYTKTEIGTQNTKINNTLKGDALATGEASIAFGSLSKAQSQAIAIGSVK 118  
DB 59 VT-----SIPQSV-----LASGLQMDVVHGTATM-----QVDGNTIIRNSVD 97  
QY 119 P-----DPN-----NGSNGVSGSHAKNESIAIGGOVLAEGDA-----SIALG 156  
DB 98 AIIWKQFNIDQNMVQFLQENNNNAVFNRTVSNQISQLKGILDSNGQVFLNPNGITIG 157  
QY 157 SDDL-----YLPKNLDLKNFEHLKHGHEILKKIQTSTDGKIKYRRTAQQGHASTAVGAM 211  
DB 158 KDAIINTNGFTASTLDISN-----NIKARN-----183

212 SYAQGHESNAFCYATAEAAYSILAVGLAAQATKQSSIAVGSNAKANAFATAIGNT--- 268  
184 -----FTEQTKDKA-----LAEIVNHGLITVKGSSYN-----LIGGKVAKNE 221  
QY 269 -VVNL-GRGVALGFGSGLDRDNNTDASAVPLGKTLADQYKATQCGDSTDFISI-GNSN 325  
DB 222 GVISVNGSGISLLAGOKITISDIINPTIY-----STAAPENEAVALNIG---DIFAKGNNIN 274  
QY 326 NNNSSIRKIIINVAGSRDTDVAVNAQLKLVEBELANRKITFKGDDNNNSV-----ERG 380  
DB 275 VRAATIRNQ-----GKLSADSVS-----KDKSGNIVLSAKEGEAEIGGVLSAQNOQA 321  
QY 381 LGNTLTITIKGDAOTNALTEANIGVVVTDGNGLK-----VKLAKELTGLTS 423  
DB 322 KGGKLMITGDKVT-----LKTGAVIDLSGKEGGETYLGSDGERGEKNGIOLAKK-----YS 372  
QY 424 VSATNKITVSNNTNNNAELOSGLTSPITGKTVDKTVISYIDGLKFTDNSNSTIATGTTT 483  
DB 373 LEKGSTINVSCK-----EKGG--FAIVG-----DIALIDGINAAGSGDIAKTG--- 415  
QY 484 ITKKKIGFAGT-----NDGVDESKPYLDNEKLKVGNSLTNSGSLTVNTTGNKQIQV 535  
DB 416 -----GEVETSGHDLFIKDNAIVDAKEWL-----LDFDNVSINAEADPLFNNTGINDFEP 465  
QY 536 G-----ANGIKFATVANNAVNTSATVGTARITTEKIGFAGTND 573  
DB 466 GTGEASDPKKNSELKTLTNTTISNYLKNAWTMITASRKLTVNSS-----INI 514  
QY 574 GVDEQAPYLDKRLKVGVRVEITTDSGINAGN-----HKTGLTNGCIANTDAVTI 622  
DB 515 GSNHL-ILHSKGORGQGVQIDGDIITSKGNLTIVSGWVDVHKRNITLDQGLNITAASV 573  
QY 623 -----KQKDKAKPTLNA-----GDGISINSNGDL-----VDSGSGNIT- 655  
DB 574 AFEQGNKAKDA---ANAKIVAQGVTVTTGEGKDPANNVSLNGTGKGLNISSVNNLTH 630  
QY 656 --TPTYNISVKTTLKNSNGTSGNNKFSVSNHNNLSVTAKDLADYLNKVNETADALS 713  
DB 631 NLSGTINISGNIT---INQTRKNTSYQTSKSHSNVSNALNL-----ET--GANFT 677  
QY 714 FKVQNGDNNAITVGKDTNGKTFNTLKLKGNGVNIITNTRATGT-VTFGIDQSNGLTTP 772  
DB 678 FIKYLSNSKGLTTOYRSSAGVNFN-----GVNG-NMSENLKEGAKVNFKLPNEMMNTS 731  
QY 773 K-----LTVGSDTNGNRLVIEQVPSADGNSTKNIIGKSLPTLPSIAPSGRNIALGNTI 826  
DB 732 KPLPIRFLANITATGGSVFFDIYANHSG-----RGAELKMSIENISNGANFTLNSHV 784  
QY 827 EEKD--KSNAAISDDVLNAGFNKNGKDKDFVSTYDVTDFIDGNATTA----- 873  
DB 785 RGDDAFKINKDLTINATNSFSLR---QTKD-----DFYDGYARNAINSTYNSILG 833  
QY 874 -TVTYDEANOTSKVAVDVNVDE-KTIELTDGCKKQGV--KTIKLETSTNGNATFTST 929  
DB 834 GNVTLGQNSSSSITGNTITTEKAANVTLEANNAPNOQNIHDRVIRKGLSLLVNGSLTGE 893  
QY 930 DDDHALVKASDIAGNLWLAEEIHTTKGTANTALTQFTTKVDENDKADDTNATITVGKDG 989  
DB 894 N-----ADIKGNL-TISESA-TFKGKTRDTL-----NITGNTNNGTAFINI 933  
QY 990 TSGKVNTLKLKGNGLDKTKDQCTVTFGINTQSLKAGDSTTLNNGN-LSIKNTASNEQ 1048  
DB 934 TQGVVKGLVNTDGDNLN-----TTHAKRNORSIIGGD--IINKKGSUNITDSNDAE 984  
QY 1049 IQVGADGVKFAWVNVGVVAGIDGTTTRITRDEIGTGTNGSLDKSKPHLSKDGGINAGKK 1108  
DB 985 IOIGN-----ISOKEGNLT-----ISDKINI-TKQ 1010  
QY 1109 ITNIOGSEIAKNSHDVATGGKIYDLKTELENKISSITAKTQNSLHEFSVADEQGNNTVS 1168  
DB 1011 IT-IRKGDIDGEDSSSDATSNANLIKT-K---ELKLTEDLSISGFNKAEITAKDGRDLTIG 1066

```
QY 1169 NPYSSYDTSKTSVITFAGENGITTKYKGVVRVIGIDTKGLTTPKLTGVGNNGKGIVIN 1228
Db 1067 N---SNDGNSGAEAKVTFFNNVDSKISADGHNVTLS-----KVTSSNG-GRESN 1115
QY 1229 SQNGQNTITGLSNPLANVTNDKGSVRTTEQGNIIKDEDKTRAASIVDVL SAGPNLQ-NG 1287
Db 1116 SDND---TGLTITAKNVEVK-DITSIKVTNITASEKVT-----TTAGSTINATNG 1162
QY 1288 EAVDFVSYDTYVNFANGNTTAKVTYDDTSKSVVYDVNVDDPTIEVKDKKLGKVTTL 1347
Db 1163 KASITTKTGDIISGTISGNTVSVSATVDLTTKSG-----SKIEAKGEANVT SATG 1212
QY 1348 TSTGTGANKFALSNGATGDALVKASDIVAHLNTLSGDIQTAKGAS-QANNSAGYVDADGN 1406
Db 1213 TIGGT-----ISGNTVNVNAGDLTVGNGAEINATEGAATLTATGN 1254
QY 1407 KVIYDSTDKYQAKNDGTVDKTEKAKDLVAQAQTPDGT LQAMNVKSVINKBQVNDAN 1466
Db 1255 TL---TTEAGSSITSTKGQVD-----LLAQ----- 1276
QY 1467 KQGINEDNAFVKGLEKASDNKTKNAAVTVGDLNAVAQTPLT FAGDTGTAKKLGTLT 1526
Db 1277 -----NGSIAGSINAA--NVTLN---TTGTLTTVA-----GSDIKATSGTLV 1313
QY 1527 IKGQTDITNKLTDNNIGVWAGTDGFTYVKLAKDLTNLSV NAGGT-KIDEKGISFVDANGQ 1585
Db 1314 INAKDAKLN-----GDASG-----DSTEVAVNASGSGSVTAATSSSVNITGD 1356
QY 1586 AKANTPVLSANGLDGGKVISNVGKGT-----KDTD-----AANYQQLNEVRNLLCLG' 1633
Db 1357 LN-----TVNGLN-----IISKDRNTVRLRGKEIEVKYIQPGVASVEEVIEAKRVLEKV 1406
QY 1634 NDNADGNQVNIADIKDPNSGSSNRVTI KAGTVLGGKGNNDTEK LATGGVQGVGVDKDN 1693
Db 1407 KDLSEERETLAKL-----GVSAREVEPNNTITVNTQNEFTTR-PSSQVLIISGRACF 1459
QY 1694 ANGDLSNVWVKTKQDG 1709
Db 1460 SSGNGARVCTNVADDG 1475
```

Search completed: September 13, 2001, 12:38:28  
Job time: 48224 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.  
OM protein - protein search, using sw model  
Run on: September 13, 2001, 12:39:59 ; Search time 76.31 Seconds  
(without alignments)  
2049.356 Million cell updates/sec  
Title: US-09-361-619-9  
Perfect score: 10356  
Sequence: 1 MNHIYKVFKNKATGFMAVA.....NGSADTOGHVGAAGAGHFH 2053  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match %	Length	ID	Description
1	1191.5	11.5	2059	2 D82671	surface protein XF
2	971.5	9.4	1588	2 A86036	probable adhesin z
3	704.5	6.8	1190	2 A82615	surface protein XF
4	623.5	6.0	5188	2 B85547	probable RTX famil
5	608	5.9	2249	2 A41477	190K surface antig
6	591.5	5.7	2240	2 B71704	cell surface antig
7	553	5.3	2514	2 F81045	hemagglutinin/hemo
8	542.5	5.2	2660	2 E85822	probable invasiv z
9	540.5	5.2	4919	2 T31105	hypothetical prote
10	526.5	5.1	2383	2 D64962	probable membrane
11	526	5.1	3029	2 S76109	hypothetical prote
12	513	5.0	4152	2 T31102	filamentous hemagg
13	512.5	4.9	2703	2 H81193	hemagglutinin/hemo
14	499.5	4.8	2020	2 C83399	ABC-type transport
15	487	4.7	5627	2 C83339	hypothetical prote
16	484.5	4.7	2893	2 A64556	toxin-like outer m
17	464.5	4.5	2154	2 F83068	hypothetical prote
18	462	4.5	2232	2 T34434	hypothetical prote
19	461.5	4.5	13288	2 T03099	mucin, submaxillar
20	460	4.4	1477	2 B43855	high-molecular-we
21	460	4.4	2468	2 A83412	hypothetical prote
22	459	4.4	2055	2 T31110	extracellular matr
23	456.5	4.4	1643	2 D71630	outer membrane pro
24	455	4.4	1336	2 A43855	high-molecular-we
25	453.5	4.4	1577	2 A35140	hemolysin A precu
26	452.5	4.4	1651	2 JC1340	outer membrane pro
27	449	4.3	2902	2 C71953	toxin-like outer m
28	446	4.3	3194	2 D71917	toxin-like outer m
29	439.5	4.2	1902	2 S06997	lactocepin (EC 3.4

30	436.5	4.2	1645	2 JN0896	crystalline surfac
31	432.5	4.2	2334	2 S32920	cell wall-associat
32	431.5	4.2	1962	2 A32634	lactocepin (EC 3.4
33	430	4.2	3535	2 E83641	probable hemagglut
34	426.5	4.1	1902	2 B45764	lactocepin (EC 3.4
35	422.5	4.1	2204	2 A70524	probable PpE prote
36	420.5	4.1	1403	2 T17372	plasma membrane-as
37	419	4.0	1965	2 S75200	fat protein - Syne
38	417	4.0	4558	2 C82199	RTX toxin RtxA VC1
39	416	4.0	2399	2 H71879	toxin-like outer m
40	416	4.0	3282	2 E82750	hemagglutinin-like
41	414	4.0	1029	2 T30852	outer membrane pro
42	413.5	4.0	2508	2 S61441	surface-associated
43	413	4.0	1459	2 T32271	hypothetical prote
44	411	4.0	4688	2 F82885	cellulose-binding
45	410.5	4.0	1848	2 A44140	

ALIGNMENTS

RESULT 1

D82671

surface protein XF1529 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C:Accession: D82671

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: D82671

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2059 <SIM>

A:Cross-references: GB:AE003982; GB:AE003849; NID:g9106554; PIDN:AAF84338.1; GSPDB:GN

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv

M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF1529

Query Match 11.5%; Score 1191.5; DB 2; Length 2059;

Best Local Similarity 23.4%; Pred. No. 6.9e-38;

Matches 548; Conservative 362; Mismatches 771; Indels 657; Gaps 101;

QY 19 VARYAKSHSTGGSCATQGVSGVRLTSFARIALAVLIGATLNGSAYAOITTKIEIGQ 78

Db 76 VTGYARFFGSPG---TAAEQQAGSRNL-----TIGGSY----VNSGQGV 114

QY 79 TNKINFTLKGDALATGASIAFGS---LSKAQGSOAIAIGSKVPDPNNGSNVNGSHAKG 135

Db 115 NDVLNKTY-----STMGSVITMNTVAGTNAIAIGSAQSSAADALKASLATKASG 164

QY 136 NESTATGGDVLAEQDASIAIGSDOLYLPKNLDLKNFPHKLTGHGHEILKIKTQSTPDGIKY 195

Db 165 ARAIAIGAKASADGVDVRLGS----- 186

QY 196 RRTAQGHASTAVGAMSYAGOHFSNAGTGTATATAAAYSLAVGLAAQATKOSSTAVGSKNAK 255

Thu Sep 13 14:19:27 2001

Db	1090	NGSOLNAVQVQASQPVYFTG-----NEGAVKRSIGOSWIS-----GESSTAGTY	1134
Qy	1186	AGENGITTKVKNQVVRVIGDQKGLTTPKL--TVGNNGKGVINSQNGQNTITGLSNTL	1243
Db	1135	SGGN-LKSVVDEAAGRIHQLA---DSPKFGVNVINNGK-----ISCVT---	1175
Qy	1244	ANVTNDKGSVRTTEQGNIIKDEDKTRAASIVDVLVSAGFNQNGENAVDPVSYDYDVNFAN	1303
Db	1176	-----AGTEETDAV---NFSOLKSISTAVDQGWLTASGANGSKVASGGTVDLKN	1222
Qy	1304	--GNTTAKVYDDTSKYTSKVYDVNDV-----DITIEVDKDKLGVKVTTLTSTGTG	1353
Db	1223	TDGNLTISK-----SGDSNDVYFNLSKDFKVGMTSGTTVWND---GVKVGSDVALGTT	1274
Qy	1354	ANKFALSNOATGDALVAKASDIVAHLNLTLSGDIQPAKASQANNSAGVYDADGNKYIDST	1413
Db	1275	GLTIANGPAVTASGIDAGSKVISHV-----ACAVSETSTDAVNGSOLNAVQVQAS	1325
Qy	1414	DNKYQAKNDGTVDKTEVAKDKLVAQAOTPDGTLAQMNVKSVINKE-----QVNDANK	1467
Db	1326	QPVYFTG-NEGAVKRS--LGOSVVISGESSTAGTYSGNLSKVYDEAAAGTILHQLADSPK	1382
Qy	1468	KOGINEDNAFVKGLEKAASDNKTNAAVTVGDLNAVQAOTPLTFAGDTGTTAKKLGELTI	1527
Db	1383	PGNVINNG---GKISGVTAGTEETDAVNFSQLSKI-----STAVDQGWTLTA	1427
Qy	1528	KG-----GQDNTNKLTDNNIGVYAGTGDFTV--KLAKDLTNLSVYNAGGTKIDEKGI	1577
Db	1428	SGANGSKVASGGTVDLKNTDGNLTISKGDSNDVYFNLSKDFKVGMTSGTTVWND	1486
Qy	1578	S-----FVD-----ANGQAKANTPV-----	1592
Db	1487	KVSNVLLDUSNELVITSHSSTSVKTLANGSVNRTVYVNGVNDVVDVVDVVDLGLSIV	1546
Qy	1593	-----LSANGLDLGGKVISWVGKTKTDIDANVQOLNEVRNLLGLG-----NMDAGNQV--	1642
Db	1547	GGASLTLSGINAGSHKITVNTAGTEDIDAVNFSQLKSVSEAVDKGWTLTASGANGSKVVS	1606
Qy	1643	NIADIKK-DPN---SGSSSNRTVI-----KAGTVLGGKGNNDTEKLTATCGVQGVODK	1690
Db	1607	GCTVDLANTDGNLAIKSGDSNDVFNLSKDFKVDVETAG-----NTVYVNTDGVKVGSDV	1661
Qy	1691	DGNANG-----DLSNVWVKTKDGSKKALLATYNAAGTNYVTNNPA---EADIRINEQGI	1743
Db	1662	SIGAMGLFIANGPSVTASGPNAGDK---VISHVAVGMADTDVAVNSQLKQAVQSVTVKAT	1718
Qy	1744	RFEHVNDGNQPVVQGRNGIDSSASGKHSAIGAQAADGEAAVAIGRQTOA--GNOSIAL	1802
Db	1719	RYSTNDGG---TQGGNYDGDGATGSKAIAAGVGTQASGEGAAVGSAGAAASGKGSTAI	1774
Qy	1803	GDNAQATGQDSIAIGTG-----NVVAGKHSG-----AIGDPSTVKADNSYSVC	1845
Db	1775	GRNAIASADGVALGDGAKDGGRAESYTGKSGVQNNVTGTVSVGDAKAGETRSISNVA	1834
Qy	1846	NNQFTDAT-----QT-----DVEGV--GNNTVVTES---NSVAL	1875
Db	1835	DAKEADAVNLQDLDAVQAQSNLQDDMRHETNINTEYFKITKDGSSASSVKGMCVNMAH	1894
Qy	1876	GNSAISAGTHAGTQAKSDGTAGTTTTAGATGTVKFGAGOTAVGAVSVGASGAEIRIQN	1935
Db	1895	GTNAAVS-GTESVALGKNTNVSADNVAIG--NGSVADRA-----NSVSGSGGSEROVTN	1947
Qy	1936	VAAGEVSATSTDAVNGSOLYKATQSIANATNELDHRHONENKANAGISSAMAMASHPQA	1995
Db	1948	VAAG---TADTDAVNVSQL---NOGLITAKQYTDGMVGNLRRSTSGGVAATAATANLQQA	2001
Qy	1996	YIFGRSVMTGTHNGQGNVAVGLSKLSDNGQWFKINGSADTQGHVGAAGVAGFHF	2053
Db	2002	YVQCRGTSYGVSSYQGOAIAVGVSAVSESHVFKPFGSANTRSHVGVGAGVGYQW	2059
RESULT	2		
AB6036			

Db	187	-----GATA-----GT-----GASSIAGLNASAV-NGAVAVGGGAL	217
Qy	256	ANA-FAATAIGNTVNLGRGVALFGQSILDRDNTDASAVPLGKTLADQYKATROGD	314
Db	218	VTPDGAVALGNSVASTGKLS---GYDPKTKTSTDSASA-----AWKSTLAA--	263
Qy	315	STDIFSIGNNNSSIRRKIIINVAGSRDQDANVNAQLKVELANRKITFKGDCDNN	374
Db	264	-----VSGDSSTNLK-TROLSGLAAGTSNTDAVNAQLKVDVDEIASR-----	306
Qy	375	NSVERGLNTLTKDQATNALTANTGAVVYDGNLKVLAKELTGLTSVSNKTIIVSN	434
Db	307	-----GWNLTASG-----ANSGNAPGSSVDLKNKNTDKNLTITKAIGSDVQF--	348
Qy	435	TNNNAELQSGGLTFSPITGKTOKTVYSIDGLKFTNDSNSIATKGTTRITKKKIGFAGT	494
Db	349	--NLNKDKV-----VTTLAVGDALLNTDGIAGLTDV-SLSTTGLA-ITD---GPVAT	393
Qy	495	NGVDSEKPYLDNEKLVGNSTLNSGLTVNNTGNKQIQGVANGIKFATVANNVANTS	554
Db	394	ASGIDAGSKVISH-----VAAGAVSETSDAVNGSOLNAVQVQAS--QPVYFTGNEGAVKR	447
Qy	555	TVGTARITBEKIGFAGTNDG-----VDEQAPYLDKERLKVGRVEIITDGINAGN---	604
Db	448	SLGQSVVISGESSTAGTYSGNLKSVDYDEA-----GRIHQLADSPKFGNVVIN	497
Qy	605	--HKITGLINGIANTDAVTIKOLKDAKPTLNAAGDGISINSNNG-----DLVDSGN	653
Db	498	NGKISGVTAGTEETDAVNFSQLKSISTAVDQGWTLTASGNSKVASGGTVDLKNTDGN	557
Qy	654	ITPTYINISVKTTLNLSNGTSGNKKFVS-----NAHDNNSLVTAKOLADLYN	701
Db	558	LT-----ISKGDSNDVFNLSKDFKVDGMTSGTTVWVNDGVKVGSDVALGTT	605
Qy	702	KVNETASALPSFKVQNGDN--SNNALTGKDT-----NGKFTNLKLGKNGVNIITNR	754
Db	606	GLTITDGPVATASGIDAGSKVISHVAVGVSETSDAVNGSOLNAVQVQASQPVYFTGNE	665
Qy	755	ATGVTEGIDQSNGLTTPKLTGVGSDTNGN-RLVIEQ-----VPSADGN	796
Db	666	--GAVKRSLGQSVVISGESSTAGTYSGNLSKVYDEAAGRIHQLADSPKFGNVINNG	723
Qy	797	STKNIKGLST-----LPSIAS-----PSGRNIALNGTTEKDKSNAASI	837
Db	724	KISGVTAGTEETDAVNFSQLKSISTAVDQGWTLTASGNSKVASGGTVDLKNTDGNLT	783
Qy	838	D-----DVLNAGFNLKN-----GKDKDFVSTYDVTDFIDGNATTAT	874
Db	784	SKSGDSNDVFNLSKDFKVDGMTSGTTVWVNDGVKVGSDVALGTTGLTI--TDGPVATAS	841
Qy	875	-----VTYDEANOTSKVAYD-VNVDK-----KTIELFGDNG--KKQLGVKVTIK	914
Db	842	GIDAGSKVISHVAVGVSETSDAVNGSOLNAVQVQASQPVYFTGNEGAVKRSIGOSWVI	901
Qy	915	LTETSTNGNATTFSTDDHALVKASDIAGNLMTAE-----IH-----	953
Db	902	SGESSTAG---TVS-----GGLNLSKVYDEAAGRIHQLADSPKFGNVINNG	945
Qy	954	-----TTKGTANTALOTFTVKVDENDKADDTNATVKGKDTSGKVNLTCLKKNGLDIK	1008
Db	946	GKISGVTAGTEETDAVNFSQLK-----SISTAVDQGWTLTASGANGSKVASGGTVDLK	998
Qy	1009	TDKDGTVTFGINTQSLKAGDSTTLNNGLSIKNTASNEQIOVGADGVKFAWNVGVGA	1068
Db	999	NTDGNLTIS-----KSGDSNDV-----VFNLSKDFKVDGMTSGT--TVVNDGVKV	1042
Qy	1069	GID--GTTTRITREIGFTGNSLDKSPHLSKDGIDAGGKKTINOSGETAKNSHAY	1125
Db	1043	GSDVALGTT-----GLTIANG-----PAVTASGIDAGSKVISHVAAAGVSETSDAV	1089
Qy	1126	TGGKIYDLKTELENKISSTAKTQNSLHEFSVADEQGNFTVSNPYSSSYDTSKTSDVITF	1185

probable adhesin Z5029 [imported] - *Escherichia coli* (strain O157:H7)  
C:Species: *Escherichia coli*  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: A86036  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: A86036  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1598 <STO>  
A:Cross-references: GB:AE005174; NID:gl2518349; PTDN:AAG58749.1; UWGP:Z50  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z5029

	Query Match	9.4%;	Score 971.5;	DB 2:	Length 1588;
	Best Local Similarity	22.4%;	Pred. No. 1.2e-29;		
	Matches	495;	Conservative 281;	Mismatches 651;	Indels 787; Gaps
Qy	1	MNHIIKVIIFNKATCTPMVAEYAKSHSTGGGSCATQQGVSRTLSPARAAALAVLVIGAT	60		
Dd	1	:     :   :   :   :   :   :   :   :   :   :   :   :   :   :			
Qy	1	MNKFVIMWPATCNVTYTSETAKRCK-----KGRSKLLISALVAGGMSSFGAL	52		
Dd	1	:     :     :     :     :     :     :     :     :     :     :     :     :     :     :     :     :			
Qy	61	LN-GSAVAQOITTK-----TEIGOTNNKINTL--KGDAIATGEASIAFGLSRAOGS	109		
Dd	53	ANAGNDNGCGVDYGSAGDCGVAICGKAKANTFMWTSGSSTAGVYDATABEOYSYAIGS	112		
Qy	110	QAIAIGSVKPPNNGSNVGHSHAKNESTAIIGDDLAEGDSIAIGSDDLVLPKNLIDLK	169		
Dd	113	K-----THAIGASMAFCVSALSEGDRSIALGASSYSLGOYSMAL	152		
Qy	170	NEPHKLHGHEILLKIQTSTDGKIKRYRTPAQGHAHTAVGAMSYAOCH-----FSNAF	222		
Dd	153	GRYSKAL-----GKLSI-----AMGDSSKAECANAIALGNATKATEINMSIAL	194		
Qy	223	GTYATAEAAYSLAVGLAAOAATKOSSIVAGSNAKANAFATAIGNTVWNLGRVALGFSGS	282		
Dd	195	GD TANASKAYSMAI LGCASSVASEBENA TAIGETEA - AENATAIGNAKAKGTNSMAMCFGS	253		
Qy	283	QILDRDNNTDASAVYPILG----KTLDQYKATROG---DSTDIFSIGN-----	323		
Dd	254	-LADKVNT-----IALGNSQALADNAIYGQGNKADGVDAIALGNGSQSRGLNTIALG	306		
Qy	324	-----SNNN-----NSSTRKLIINVAG---S	342		
Dd	307	TASNATGDKSLALGNSSSANGINSVALGADSIADLDTNVSVGNSLKRRIKRVNYKNCAIGKS	366		
Qy	343	RDTADVNVVAOLKLVEELANKKITFKGDDNNSVERGLCNTLTIKGDAQTNALTEANTIG	402		
Dd	367	DSYDAIINGSQLYAI-----SSVAKRLGGGAADVDDGTVTAPTYNL--	408		
Qy	403	VVTDGNGLKVLAKEI TGLTSVSNATKNITVSNP---NNNNAELQSGLTFSPITGTKDK	459		
Dd	409	-----KNGSK-----NNVGAALVDLENTLWDQTKGYKSYAHGTSPTASVIID-	453		
Qy	460	TVYSIDGLKFTDNSNSTATKGTTRITRKKKIGFPAGTNDGVDESKPYLDNKEKLKVCNSTLWS	519		
Dd	454	-----VADGTITAS-----SKDAVNGSQLKATNDVDEA	481		
Qy	520	GSLTVNNTTGNGKIQVGANGIKATVANNVANTSATVGTARI-----TEEKIGFACTNDG	574		
Dd	482	NTANIAIYTN-----IATNTANIATNTNTNIUTSDVSGDALLWNETTKFAAHH--G	536		
Qy	575	VDEQAPYLDKERLKVGRVEITTDOSGINAGNHKITGLTNGTIANTDAVTIKOLDAKPTLNA	634		
Dd	537	QDTTSKITN-----VKDADLTADS-----TDVNGSQLK-----TT	567		
Qy	635	GDGISINSNGDGLVDSSGNTTTPPYNISVKTTKLNSNGTSG-----NNKFSVSYAHDNN	688		

Db	568	NDAVATNTN---IANTSNIA	TNTNINL	ETVTN	LGEDAL	KWDKONGV	PTAAHGTETT	625	
Qy	689	SLVTAKDLADYLNKVN	ETAD	SALP	KFYQNGD	SNNAITV	GKDTNGTNTLKLKGNGV	748	
Db	626	SKIT----						657	
Qy	749	NITN	RATGTV	TFEGID	OSNGLT	TPKLT	VGSDT-----NGNRLVIEQVPSADGNSPKNII	802	
Db	658	TNTNTIATNTNI----						706	
Qy	803	KGLSPTLPSIASPSGR	NIALGNT	IEEK	DKSNAAS	IDDVLNAG---FNLKNNKGKDKDFVSTY		860	
Db	707	TNI-----						737	
Qy	861	DTVDFIDGNATTATV	IDEANOTS	KVAYD	VNVDEK	TIELTGDNGKKOLGV---KTIKLTE		917	
Db	738	----						759	
Qy	918	TS--TNGNA-----							
Db	760	TNYNVGDDALAAINS	SFSTSLGD	ALLMDA-TAGK	ESAK---HGTNGDASVI			806	
Qy	971	VDENDKADOTNAITV	GKCGTSK	VNTLKLKGN-----GLD	LKTDKDGTV---TFGIN			1020	
Db	807	---TDVADGEIS-----						858	
Qy	1021	TQSLKAGADSTLLNN	GLSIKNTAS	NEQIQV	GADGVK	FPANVNGV	GAGIDGTTTRIRDE	1080	
Db	859	NADYD	NVGDA-----					890	
Qy	1081	IGFTGNGSLDKSPH	LSKDGINAG	GKKITITQ	SGEIAKNSH	DAVTGCKIYDLKTELENK		1140	
Db	891	-AFSAAHGK-DKT-						936	
Qy	1141	ISSTAKTAQN---SL	HEFSVADEQGN	FTVSNP	SYSDTSK	TSDVITFAGENGITTKVNK		1197	
Db	937	LGGDAEVNADGTT	APTATY	TIANAEYN-----				978	
Qy	1198	GVVRVGIDQTKGLT	PKLTAVGN	NNKKGIVINS	ONGONTITGL	SNTLANVTN	DKGSVRTE	1257	
Db	979	-----						1008	
Qy	1258	QGNIIKDEKTRAA	SIVDLSAG	FIQNGE	AEVDFVSTY	DTVNFANGNTTTAKTYVDTS		1317	
Db	1009	-----						1029	
Qy	1318	KTSKVYDV--NV	DDTTIE-----					1367	
Db	1030	ONTOIINQ	LAGNTDATYIO	ENGAGIN	VYRTDGD	LAFNDASAO	GVGATGYNSVAKGS	1089	
Qy	1368	LV-----						1419	
Db	1090	SVALGQGSYSDVTG	TALGSSSVSR	VIAGSRDYS-----				1144	
Qy	1420	AKNDGTVDKTK	VEYAKDKL	VAQAQTPD	GTLLAOMN	VKSVINKEQVNDANK	KQGINEDNAFVK	1479	
Db	1145	ALSIG-----						1159	
Qy	1480	GLEKASDNKTKNA	AVTVGDL-----					1533	
Db	1160	-----						1214	
Qy	1534	TNKLTDNNIG	VVAGT-----					1584	
Db	1215	VN--GDKGIGIGY	GVAVDANALNG	IAIGSNAQV	IIHVNSTAT	IGNSGSTTTTRG	AQTYNTAYNM	1272	
Qy	1585	QAKANTPVL	SANGLDLG	GKVI	SNVKG	TGTD	DAANVOQLNEVRN	LLGLNDNADGNOVNI	1644
Db	1273	DAPONS	VGEFSVGSAD	QGRITN	VAAGSAD	TDAVNGQL-----		1313	
Qy	1645	ADIKDPNSGSS	NRRTVIR	KAGTV	VLGKGN	NDTEK	ATGCVQVGV	DKGNANGDLSN	1704
Db	1314	TDAQV	SQNTQSIJTN-----					1350	

Thu Sep 13 14:19:27 2001

1705 TQDGSKKALLATYNAAGQTYNNPABEADIRNEQGIREFHVDNGQEPVVGQNRGID 1764  
 1351 -----TGSTKYFKTN-----TDGVD 1366  
 1765 SSAGKHSVAIGFOAKADGANAIVAGQTOAGNOSTAIGNAQATGDOQIAIGTGNVAG 1824  
 1367 ASAGKDSVAI-----GSGSIAADN-----SVALGTGSV--- 1396  
 1825 KHSAGIDPSTVKADNSYSVGNNOFTDQTQDVFVGNNITVTSNSVALGSNSAISAG 1884  
 1397 -----ATEENTISYGSSTNORRITNVAAGKNA----- 1423  
 1885 THAGTOAK-KSDGTAGTTTATAGTGVKGFAGOTAVGASVAGSAGAEERRIONVAAGEVSA 1943  
 1424 TDAVNAQLKSEAGGVRYDTKAGSID--YSNITLG---GNGGTTTRISNVASG---V 1474  
 1944 TSTDVANGSOLYKATQSTANATN---ELDHRIHQENKANAGISSAMAMASMPQAYIPG 1999  
 1475 NNNDVYNAQLKQSVQETKQYTDQRMVEMDNKLSKTESKLSGGIASAMAMTGLPQAYTPG 1534  
 2000 RSMVTGGTATHNGQAVAGLSKSDNGOWEKFINGADTQCHVGAAGVAGGFHF 2053  
 1535 ASMASIGGGTYNGESAVALGVSWNSANGRWYKLOGSTNSOGESYSAALGAGIQW 1588

RESULT 3  
 A:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: A82615  
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: A82615  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1190 <SIM>  
 A:Cross-references: GB:AE004017; GB:AE003849; NID:99107083; PIDN:AAF84783.1; GSPDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to Genbank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Franco, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins,  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF1981

Query Match 6.8%; Score 704.5; DB 2; Length 1190;  
 Best Local Similarity 22.9%; Pred. No. 1.2e-19;  
 Matches 360; Conservative 179; Mismatches 508; Indels 525; Gaps 65;

QY 624 QLKDAKPTLNAGDGSINSNNGDLVDSSGNIPTPTNYI-----SVKTKLNSNGT 673  
 Db 2 QIHTALPMVRLGG-----IEGVDPFLPKYKIGQSLQHAHWITSAASSKKGTOPR 50  
 QY 674 SGNKFSYSNA---HNNSL--VTAKDLA---DYLKNVETADSALPSFKVONGDSNNA 725  
 Db 51 RSNAMTAKRSKLRDRRLHVLVLAASTGYTGKVAQV-----YVNSDSTENC 102

Db 868 -----GATGTRSIAGVGTLASAEGATAVGSG 894  
 QY 1793 TQA-GNQSTAIGNAQTGDSIAICTG-----NVVAGKHS-----AIGDPS 1834  
 Db 895 AAASRGSTAGRNAVASAGSVALGDGARGAESYTGKYGQNNVTGTVSVGDAS 954  
 QY 1835 TVKADNSYSVGNNOFTDAT---QTDVFG-----VGNNI-----TVTESNSVALGSN 1878  
 Db 955 KGETRTVSNAADKEATDANLRQLDRVAQDANRYVDNKIESLSEBGTFFVKVNSL---NN 1011  
 QY 1879 SA--ISAGTHA-----GTOAKKSDGTA-GTTTTAGATGVKVGAGATAVG-----A 1921  
 Db 1012 SATPIAAGVDATAIGVGATASGADSTAMGNKASASADNAV-----AIGNHSHVADRANT 1064  
 QY 1922 VSVGASGABERRTONVAAGEVSTATSDAVNGSOLYKATQSIANATNELDRIHONENKANA 1981  
 Db 1065 VSVGSAGSERQVTNVAAG---TADTDAVNVSQ---NOGLITAKQYTDGVGVGLRRDITDG 1118  
 QY 1982 GISSAMAMASMPQAYTPGRSMVTGGTATHNGOGAVAGLSKLSDNGOWVFKINGSADTQG 2041  
 Db 1119 GVAALATATANLPQAYTPGRGMTSVGSSYRGOSAIJAVGVSSVSESGRWVFKFSGSANTRS 1178  
 QY 2042 HVGAAYGAGGFH 2053  
 Db 1179 QVGIGAGVGVQW 1190  
 RESULT 4  
 B85547  
 probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7)  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
 C:Accession: B85547  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoculis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: B85547  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-5188 <STO>  
 A:Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:206  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: 20615  
 Query Match 6.0%; Score 623.5; DB 2; Length 5188;  
 Best Local Similarity 22.6%; Pred. No. 1e-15;  
 Matches 526; Conservative 287; Mismatches 846; Indels 673; Gaps 126;  
 QY 32 SCATQGVSVRTLSFARIAALAVLVIGATLNGSAYAAQITTKIEIGQTNKNNTLKGD-- 89  
 Db 1497 SVPTGLASLTASSYTVNAS---VSDKARNASASATHNLT--VDLAAPVVTINTVAGDDI 1550  
 QY 90 --ALATGEASIAFGSLUSKAGGSQAIAGSVKPPNNGSNGVGS-----HAKGNESIA 140  
 Db 1551 INATEHQQAIIISGATGATGNTSV-----TIGTTTYYTVLDANGNWSIG 1597  
 QY 141 IGGDV---LAEGDASI-----AIGSDDLVLPKNLDLKNFHKLIH 177  
 Db 1598 VPASVISAQAQGDVITATVTSAGNSGTASHTVTVVALGAPVLAI-----NTA 1646  
 QY 178 GHEILKIQITSDGKIKYRRRAQAG--HASTAVGAMSY-----AQGHFS---NAFGTYATA 228  
 Db 1647 VDDIINAKEGADLAITGTSNQAPAGTQIIVTLNGQNYTTTADAGNWSVTVPASRVYALG 1706  
 QY 229 EAYSIAVGLAAQATKOSSTIAGVSNKAKANAFATATGGNT-----VNLGR-QVALG 279  
 Db 1707 EATYT--VTAAATDADGNSGSASHNVQVN---TALPGVTINVVATDDIINAAGAEQET 1760

QY 280 FGSQLDRONNTDASAVVPLGKTLAD-QYKATRGD---STDI-----FSIGNSN-NNNS 329  
 Db 1761 ISQGV-----TGAAGADTVTVTLGGATYATATQANLSWSVDVPASALQELNGELTISA 1814  
 QY 330 SIRRKIINVAGSRD--TDVNVAAQLKLVEELANRKITFKGDDGNNSNSVBERG----- 380  
 Db 1815 SVTNSVGTNGTGREITIDANLPGLR-VDIVA-----GDDVNIIEHQALVITGS 1864  
 QY 381 -----LGNLTUTIKGDAQT-----NALTENIGVVTGNGL 410  
 Db 1865 SSGLAAGSNVTLTINQOTYVAVALDGTWSVGPVAVDVSAMPAGSVTIAASGSTSAGNPV 1924  
 QY 411 KV--KLAKELTGL-TSVSATNKITVSTNNNAELQSGGLTFSPITGKTGKTDTKTVYSIDGL 467  
 Db 1925 SVTHPVTVDLSAVAVSINAITADDVINAEEKGAALTLSGST-----SCVEAGQTVTVTFGG 1980  
 QY 468 KFTNDSNSTATKGTTRITKKKIGFAGTNGDVDESKPYLDNEKLKVGNS-----LN 518  
 Db 1981 K--TYSATVAANGSMWTSVPAADMAALRDGASQAQSVSNV--GNSATTHAYSVDAS 2035  
 QY 519 SGSLTVNNTGKQIOVGANGIKPATVANNVANTSATVGTARITEEKIGFAGT--NDGV- 575  
 Db 2036 APTVTINTIAGDDILNAAEAGAAITITGSSTAEAGQTV--TTLNGTNTGTGTQTDGWS 2092  
 QY 576 DEQAPYLDKERLKVGRVEITTDSCINAG-----NHKITGLT-----NGIANTDAYTIK 623  
 Db 2093 SVSVPASADJSTLTASXYTVNAAVSDRAGNPASVNNHNTVDTSVPVVTVINTVAGDDVFNAT 2152  
 QY 624 QLKDAKPTLNAGDG-----ISINSNN-GDLVDSGNTTPTYNISVKTTLN--SNG 672  
 Db 2153 EHAQAQIISGSATGATGTSVTVTIGTNTTFTVLDASG-----WSVGVPSVVSALANG 2207  
 QY 673 TSGNNKFSVSNADHNNLSLVTAKDLADLYLNKVNETAASLP--SFKVQNGDNNNA----- 725  
 Db 2208 TVTINA-SVTDAGNSGSATHQ-----VTNVTGLPTITFNALSGDNLNADESKGQ 2256  
 QY 726 -ITVQKDTNCKTEN--TLKLKGENGVNITNRTATGTVTFGIDQSNGLTTPKLTVGSDTN 781  
 Db 2257 PLTISGSGTGLATGAQVTVTLNGHN-YSATTD-ASGNWT-----LTVPVSDLAALQG 2306  
 QY 782 GNRLVIEQVPSADGN---STKNIIKGLSPTLPSTASPSGRNIALGNTIEEKDSNAASID 838  
 Db 2307 ANYTVSASATSAAGNTASSQANLL--VDSGLPDV-----TI-----NTVAGD 2346  
 QY 839 DVLNAGFNKNNKDKDF-----VSTYDTPVF- IDGNATTATVYDEANOTSKVAYDVN 891  
 Db 2347 DLIINAA---EAGADQTI SGVWTRAAAGDTVTVTLGNGTYTATV-----QSNLSWSVS 2395  
 QY 892 VDEKTIELTGDNGKKQGVKTIKLTETSTNGNATTFSTDDHALVKASDIAGNLTLAE 951  
 Db 2396 VPTADLQALGN-----GDLTITASVTNANGTCSGPRD-----ITIDANLPGL--- 2438  
 QY 952 IHTTKGTANTALQTFVTKKVDENDKADDTNAITVGKDTSGKVNTLKLKGNGLDIKTDK 1011  
 Db 2439 -----RVDTVAGDDIVNSIEHQ-----ALVITGGS----- 2464  
 QY 1012 DGTVTFEINTQSGIKAGDSTTLNNGLSIKNTASNEOIQVAGDVKFAMVNNGVYAGID 1071  
 Db 2465 -----SGLNAGAVLTVTINSVAYSAT-----VQADGWSVGVIPAAVNSAWPA 2506  
 QY 1072 GTTTRITRDEIGFTGNGSLDKSKPH-----LSKDGINAGGK-----KI 1109  
 Db 2507 GPLTVEVD--GQSSANNPVSHPFTVDLPAVAISINTVASDDVINAAEKGNTJTLSGST 2564  
 QY 1110 TNIQSGEIAKNSHDVT-GGKIY-----DLKT-----ELENKISSTA 1145  
 Db 2565 SGIESQGV-----TVTFGKTYTASVAANGSWSVNVPAADLATLPESANVQASVSAS 2619  
 QY 1146 KTAQNSLHEFSVADRGNNFTVSNPYSSD---TSKTSDDVITFAG-----ENGITTKVN-K 1197  
 Db 2620 GNSASATHAYSV-DASAPTLTI-NTIASDDILNAAEAGSPLTISGTSTAGTGTGTVTTLN 2677  
 QY 1198 GYVRVGIDQTKG---LTPPKLTVGNNGKGINV---SQNGQNTITGLSLANLVNDKG 1251

A:Reference number: A41477; MUID:90354033

A:Accession: A41477

A:Molecule type: DNA

A:Residues: 1-2249 <AND>

A:Cross-references: GB:M31227; NID:g152465; PIDN:AAA26380.1; PID:g152466

A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for resi

C:Keywords: surface antigen; tandem repeat

F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>

Db 2678 CATYTGTVQADGNSVSVPTSAFGALNASNTVSATVNDKAGNPGSASHNLAVDTTAPVL 2737  
Qy 1252 SVRTTEQNIKDEKTRASIVDVLGAGNLOGGEAVDFVSTYDYNFANGNNTTAKV 1311  
Db 2738 TINTVAGDIIINDEAHOAL-VISGTSG-----GEXGDVSV-----VLNGKTYT--- 2782  
Qy 1312 TYDDTSKTSKVVDVNVDDTTIEVKDK-KLGKVTTLTSTGTGA-----NKFALSNQ 1362  
Db 2793 -----TTLDASGNWSVGPAAADVTLGSGAQTITASVDSRAGNSDD 2823  
Qy 1363 AFGDALVRASDIVAHNLTLSDG---IQTKAGCASQANNSAGYVDADGNKVIYDSDNKKYQ 1419  
Db 2824 ASRTVTVLSAPVISINITAGDDVINATEKGSDLA--LSGTSODOPAGTAIVTLVNGQNS 2881  
Qy 1420 AKNDGTVDKTEKAVDKLVAAOQTPDGTLAQMNKSVINKEQVNDANKKOGINEDNAFVK 1479  
Db 2882 ATTDSAGNSVTVP-----ASAVSALGEATYSVTASVTN-----AQ 2917  
Qy 1480 GLEKAASDNKTKNAAVTVDLNAVAQTPLTFAGDTGTTAKKLGETLTILKGGOTDTNKLTD 1539  
Db 2918 GNSSTASHNVQNTALPGITINPVATDDIINASEAGSAQTISGQVGTGAAGSTVTVEL-- 2975  
Qy 1540 NNLGVVAGTGTFTVKLAKDLTLNLSVNAAGTKIDEKGISFVDANGQAQKANTPVLISANGLD 1599  
Db 2976 -----GKTYTATVOADL-----SNVSVPAADWQAL-----NGELTVNASVTNA----- 3016  
Qy 1600 LGKGVISNKGKTKD--TDAANVQOLNEVRNLLGLGNDWAGNQ-VNIAD-1KKDPNSGS 1655  
Db 3017 -----VNGTSGTRDTIDA-----SLPGLRVDFTVAGDDVNVIIIEHAQAQVITGS 3061  
Qy 1656 SSNRTVIKAGTVLGGKNDT---EKLATGGVQVQVDKDGNDGNSLVVWVKTKQDGSKK 1712  
Db 3062 SSG--FAAGTALTVINNOTTAATVLANGNSVGPAT-----DVSN-W-----PAGTLN 3108  
Qy 1713 ALLATYNAAGQTNVVTN-----NPAAIDRIN--EQRIRF-----FHVNDG 1751  
Db 3109 ITVSGANSAGTQTSITHPLTVDLTAVAIMSNISITSDDAINAAEKGAALTLSGSTSGVEAG 3168  
Qy 1752 NOEPVVOGRGIDS--SASGKHSV---AIGFOAKADGEAAVAIGRTOAGNOSIAIGDNA 1806  
Db 3169 QTVTVTFGGTYYTTVVAANGSNWSTVPADLAALROGDASAQV-RVTNVN-----GNSA 3221  
Qy 1807 QATGDQSI-----AIGTGNVAGKHSGA---IGDPSTVKANDSVS--GNNNQF 1850  
Db 3222 TATHEYSVDSAAPTVTINTIASDNIINASEAAAGVTVSGTSTAQTGOTLTVLNGTNYOT 3281  
Qy 1851 TDATQTDVFGVGN-NITVTESNVALGNSAISAGT---HAGTQAKKSDG-TAGTT----- 1901  
Db 3282 T--VOTD-----GSWSLTPASDLTALANGYTLTATVSDLAGNLGSAKGVTVDTTAPVI 3335  
Qy 1902 ---TTAG-----ATGTVKG-----FAGQ-----TAVGAVSVGASG 1928  
Db 3336 SFNTVAGDDVINNEHIAQIISCTATGAVAGDRLVYTIAGQQVVTSTDSAGNSVGVPA 3395  
Qy 1929 AERRTONVAAGEV--SATSDAVNGS-----QLYKATOSIANATNDELDRHONENKA 1979  
Db 3396 SV--ISGLADGTVTISATITDSAGNSSTQTHNVQVNTAAVLSVSTISGNLI-----NAA 3449  
Qy 1980 NAGISSAMAMASMPQAYIPGRSMVTGGIAT--HNGOGAVAVGLSKLSDNQW 2029  
Db 3450 EAG--SALTLSG-----TGTNFATGTVTVLLNGKGYSAT-----IOSNGSW 3489

RESULT 5  
A41477  
190K surface antigen precursor - Rickettsia rickettsii  
C:Species: Rickettsia rickettsii  
C>Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text\_change 08-Oct-1999  
C:Accession: A41477  
R:Anderson, B.E.; McDonald, G.A.; Jones, D.C.; Regnery, R.L.  
Infect. Immun. 58, 2760-2769, 1990  
A:Title: A protective protein antigen of Rickettsia rickettsii has tandemly repeated, nd

Query Match 5.9%; Score 608; DB 2; Length 2249;  
Best Local Similarity 22.5%; Pred No. 1,3e-15;  
Matches 491; Conservative 224; Mismatches 765; Indels 704; Gaps 114;  
Qy 82 INNTLKDALATGEASIAFGSLSKAQSQAIAIGSVKPPDPNNGSNVNGSH----- 132  
Db 13 IQOGLKAALFTTSTAIMLESS-----SGALGVATGVIATNNNAFNNYGNWNWEITAAG 68  
Qy 133 -----AKNVESTAIGDVLAEGBASIAI-----GSDDLVLPKNLDLKNFEHKLII 176  
Db 69 VANGTPAGGPQNMAFTYGGDYTVTADAADRIKAINVAGT-----TPVGLNI----- 116  
Qy 177 HGHEILKKIQTSTDGKIKYRR-----TRAQGHASTAVGAMSYAQGHFSNA-----FGT 224  
Db 117 -----TQNTVVGSIITKGNLLPVTLNAGKSLTLNGNNAVAANHGFADPADNYTGLGN 168  
Qy 225 YATAEAAYSLAVGLAQATKQSSIAVGSN-----AKANAFATAAIGNT-----VNL 272  
Db 169 IALGGANAALII-----QSAAPSKITLAGNIDGGIITVKTDA-AINGTIGNTNALATVNV 223  
Qy 273 GRGVALGFGSOILLDRNN-TDASAYVPL---GKTLADQYKATROGSDTDFSI----- 321  
Db 224 GAGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQV 283  
Qy 322 -GNSNNNSSTIRKLIINVAGSRDTD--AVNVAQLKLVLELANRKIT-----PKGDGDNNS 374  
Db 284 TCDIGTNS---LATISVGAGTATLGGAVIKATTTKLTDAASAVKFTNPVVVTCADNTG 340  
Qy 375 NSVERGL---GNLTITKGAQTNALTEANIG---VVTDCNGLKVKLAKELTGLTSVSAT 427  
Db 341 NA-NNGIVTFGNTSVTVGNVGNATLATVNVGAGLQVGVGVKANTINLTNDSAVTFT 399  
Qy 428 NKITVNTNNNAELQSGGLTF---SPITGKTDKTVYSIDGLKFTNDSNSIATKG--- 480  
Db 400 NPVVVTCAGIDNTGNANNGIVTTCNSTVTG-----DIGNTNALATVNVGAGTATLGGAVI 454  
Qy 481 ---TTRITTKKIGFAGTNDGVDESKPYLDNEKLVGNSLTNLSGSLATVNTTGNKQIQV-G 536  
Db 455 KATTTKLTNAASVLTLTN-----ANAVLTG---AIDNTTGGDNVGVNLN 494  
Qy 537 ANGKIFATVANNVANT-----SATVG-----TARITEEKIGFAGTNDGVDEQ 578  
Db 495 LNG-ALSOQVTIGNTNSLATISVGAGTATLGGAVIKATTTKLTDAASAVKFTNPVV--- 550  
Qy 579 APYLDKERLKYRVEIT--TDSGI--NAGNHKIT---GLTNIGIAN----- 616  
Db 551 -----VTCAIDNTGNANNGIVTFTGNSTVTDIGNTNSLATISVGAGTATLGGAVI 601  
Qy 617 -----TDAVTIKOLKDAKPTL-----NAGDGISNSNNGDLVDSGGNI--TTPYN 660  
Db 602 KATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQVTDGIGNTNSLAT 661  
Qy 661 ISVKTTKLNSGTSGNNKFSVSNADHNNSLVTAKLADLYLNKRVNETADSAALPSEKVGND 720  
Db 662 ISVGA---GTATLGGAVIKATTTKITNAVASVK---FTNPVVVTG-----AIDSTG 706  
Qy 721 NSNNAI---TVGKDTNGKTFNTLKLKGENGVNITTNRATGTVTTFGIDQSNQ-----LTPPK 773  
Db 707 NANGIVTFGNTSVTVTDIGNTNAL-----ATNVGAGTATLG-----GAVIKATTK 754  
Qy 774 LTVGSD-----TNGNRLVIEQVPSADGNSTKNIIK-----GLSPTLPSPASPSG 817  
Db 755 LTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQVTDGIGNTNSLATISVGAG 814

QY 818 RNALGNTIEKDK--SNAASIDVNLNAGFNKNNKDKDFSTYDTVFIDGNATTATV 875  
Db 815 TATLGAVIKATTTKLTNAASVLTNANAVLTG--AVDNTTGGDNVGVNLNGLSQQV 871  
QY 876 TYDEANOTSKVAVDVNDEKTIELTDGNGKKQKQGVKTIKLE-----TSNGNATTFSTD 930  
Db 872 TGDIGNTSLATISVGAGTATL-----GGAVIKATTTKLTNAASVLTNANAVLTGAI 925  
QY 931 D-----DHALVKASDIAGNLTALAEIHTTKGTA---NTALQTFVKKVDEN 974  
Db 926 DNTTGGDNVGVNLNGLSQTVDIGNTSLA--TISVGAGTATLGAVIKATTTKLTDA 984  
QY 975 DKADDGNALTY--GKGTGSKVNTKLKNGKGLDKDKDGTVTFTGNTOSGLKAGDSTTL 1033  
Db 985 SAVKFTNPVVVTGAIDNTGNAN-----NGIVTF-----TGNSTVT 1019  
QY 1034 NNGLSIIKNTASNEQIOVGADGVKFAWNVGVVGGAGIDGTTTRTRDEIGFTGTN-----G 1088  
Db 1020 GNVG-----NTNALATVNVGAG-----LLOVGGVVKA---NTINLTDNASAVFTNPVVVTG 1069  
QY 1089 SLDKSPHLSKDGINAGGKKITNIQSGEIAKNSHDAVTTGGKIYDLKTELENKILSSTAKTA 1148  
Db 1070 AID-----NTG-----NANGIYVFTTGNSTVTG-----NVGNT----- 1097  
QY 1149 QNSLHEFSVA-----DEQGNFTVSNPYSSYDTSKTSVDITFAGENGITTKVKNGVVVRVGI 1204  
Db 1098 -NALATVNVGAGLQVGGVVKA---NTINLTDNASAVTF-----TNPVVVTGAI 1143  
QY 1205 DQPKGLTTPKLTVGN--NNGKGIWINSQONTITG--LSNTLANVNTDKGSVRTEOGNII 1262  
Db 1144 DNT-----GNAING-----IVFTTGNSTVTDIGNTALATVNVGAGITLQAGSL 1189  
QY 1263 KDEDKTRAASIVD-----VLSAGFNQLQNGEAVDFYSTVDTVNFANGNT----- 1306  
Db 1190 -----AANNIDFGARSTLEFNGPLDGGKAIPY---YFKGAIANGNAILNVNTKLLT 1239  
QY 1307 -----TTAKVYDDTSKTSVVYDVNVDDTTI-----EVKKKLGKVTTLTSTGTGA 1354  
Db 1240 ASHLTGTVAEINIGAGNLTIDASVGDVITLNAQINFRARSDSVLVSML-----TCGVV 1295  
QY 1355 NKFPALNQATGDALVADSIADVAHLNTLSGDIQT-----AKGASOANNAGYV-----DAD 1404  
Db 1296 NNTLL-----AADLVAP-----GADEGTVFVNGVNGVNLVNGSVNAGTARNITGDGG 1340  
QY 1405 GNK-----VIYDS-----TDNKYYQAKNDGTVDKTKEVAK-----DKLVA 1439  
Db 1341 GNKFTLLIYNNAVITDDVNLGIONVLINKNADFTSSAFNAGAIQINDATYITIDANN 1400  
QY 1440 QAOTPDGTL--AQMNVKSVINKEQVNDANKKQGINEDNAPVKGLEKAAADNKTNAATV 1497  
Db 1401 NLMIPAGNIQFAHADAQLVLQNSSGNDRTITLGANTID-----PDNDDEGIVI-- 1447  
QY 1498 GDNAV-AQTPLTFAGDTGT--AKKLGELTIK-GQDNTNKLTDNNIGVVAGTGDFT 1552  
Db 1448 --LNSWTACKKLLTIAG--GKTFGGAHL--QTLFLKAGDCST-----AGTFNT 1491  
QY 1553 VKLAKDITNLNSVNAVGGTKIDKIGISFVDANGQAKANTPVLSANGLDLGGKVISNVGKGT 1612  
Db 1492 TNIVLDIT-----GQ-----LELG----- 1505  
QY 1613 KOTDAANV-----QQLNEVRNLLGLGNAD-----GNQVNIADIKDPNPGSSSNRT 1660  
Db 1506 --ATTANVLFNDVAVQLQTGTNTGGFLDFPNKNGMVTLLNNVNVAGAVQ--NTGGTNGT 1561  
QY 1661 VIKAGTVLGGKGNDEKLA---TGGVQGVQDKDGNAN--GDLSNVVWVTKQDKSKALLA 1716  
Db 1562 LI-----VLGASLNVRVNGIAMLKVAGNVTIAGKKVKGIEIGTGTNT-----ITLPA 1611  
QY 1717 TYNAAGOTNVVTNPNPAEADIRINEQIGIRFHVNDGNOEPVQVQNRNGIDSSASKHSVAIG 1776  
Db 1612 HFNLTG-----SINKTGGQALKLNFNMNGSGVGV----- 1641

QY 1777 FQAKAGEAAVATGROTAQGNQSIATGDNAAQT-----GDQSI-----IGTGNVYVAGKHGA 1829  
Db 1642 -----GTAANSVCDITTAGATSFASSVNAKGTATLGGTTSFANTFTNTGAVTLAK----- 1691  
QY 1830 IGPSTVVKADNSVSVGNNOFTDATQTDVFCVGNNTIVTESNSVALGCSNALSAGTHAGT 1889  
Db 1692 -----GSITSFAKNVTATSF--VANSATINFSNLSAFNSN--ITGG----- 1728  
QY 1890 OAKKSDGTAGTTTTTAGA-----TCTVKGFAGQTAAGVAVSVGA--SGAERRIQNVAAGEVS 1942  
Db 1729 -----GTTLTLCANQVYTTGT--GSFDTLTLTNTTFDGAAGSGGNILIKSGTLDLS 1778  
QY 1943 ATSTDVAVGSQLYKATQSIANATN 1966  
Db 1779 GVSTLAL-----VVVTATN 1791  
RESULT 6  
B71704  
cell surface antigen (sca3) RP451 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
C:Accession: B71704  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499  
A:Accession: B71704  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2340 <AND>  
A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAAL4908.1; PID:g3886  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: sca3; RP451  
Query Match 5.7%; Score 591.5; DB 2; Length 2340;  
Best Local Similarity 21.5%; Pred. No. 5.9e-15;  
Matches 534; Conservative 307; Mismatches 839; Indels 805; Gaps 139;  
QY 35 TGOVGSVRTLSFARI---AALAVLVIGATLNGSAYAAQIITKIEIGOTKINNTLKGDAL 91  
Db 123 TSIITGVGTGTTFNIDFAGKNATLQINKDLN-----ITTKID-----NTVAG--- 164  
QY 92 ATGEASIAFGSLSKAQSQAIGSVKPDPNNGS-----NGNVGSHAKGNESIAIGDVL 146  
Db 165 -----NNGSITFEGSIIISNHIGYTNSSL--TLCDGNI----- 230  
QY 147 AEGDASTAIGSDDLYLKP--NLDLKNFHLIHGHEILKKIQTSTDGKIKYRRTAQGHA 204  
Db 193 GNCEAKI-----YAPEANNITINAKNINLTNNLSL---TLCDGNI----- 230  
QY 205 STAVGAMSYAQGHFSNAGFYATAAAYSLAVGLAAQATQSSIA-VGSNAKANAFATA 263  
Db 231 -----TTLKGNNIT--TEIDGQGLNLAYDLGSSIIITGIDIGTGLDITNVLLGSA 281  
QY 264 IGGNTVNLGRVAGFGSGOILDRDNNNTDASAVVPLGKTLADQVKATROGSDSTDFISGN 323  
Db 282 TFSNTILK-ATNINLKHNTSTLNDN-----IIVIGN 313  
QY 324 SNNNSNSIRKIIN--VGAGSRDPTDAVNVAAQLKVEELANRKITTFKGDGNNNSN----- 375  
Db 314 IKGNNN---KDILNFKVHGTNLDNEMIIPAPQ-----TGTNLFKNATNGNINLNI 365  
QY 376 -SVERGLNLTITKGDQATNALTEANTGVVTDG-----NGL-----KVKLAKELTGLTSV 424  
Db 366 LKFSGGHGKTLNQGKTKVDNLVFPAD--SVLDSGTISYVGLDLDCTVTFNNSNVNGGLII 424  
QY 425 SATNKITVSTNNNNELQ--SGGLTFES--PITGKT-----DKTVYSTD-----GLKFTNDS 473  
Db 425 NAKNTISAKLLNATKAKIQINANLTMNHPSGDISDRIADNTIITIDAKNGNVNLLNN 484



Thu Sep 13 14:19:27 2001

474 NSIATKG-----TTRIKKK-----IGFAGTND-GV-----DESKPY 504  
 485 AKIIFEGADSMALINTGVADTRFTTYNNLNSGNDYGVIVKIEAKKVIITANOSGPY 544  
 505 L---DN-EKLA-----VGNSTLNSGLRVNNT---TGNKOIOVGANG-IRF-----542  
 545 TIGQDNTIRKELIVEGAGDIIIDTFTKLLSINSTGQITENRTLDLGGAGNIAFGKHG 604  
 543 ATVANNVANTSA-----TVCTARITEKIGFAGTNCVDEQAPYLDKERLKVGRVE 593  
 605 TLVVGVTGSIITSENNOGILITINSNIT---GVIGTNE-----IGLKLNVIGADP 652  
 594 ITDSDGINAGNHKIGTLN-----GIANTDAVT-----IKQ 624  
 653 VTCSANVFAS-----VALNPPSVLLIADGVTLTGEVTHNNTKGVLSIGTGNITGQIGT 708  
 625 LKDAKPTLNAGDGIS-INSN-----NGDLVDSGNTTTP-----657  
 709 NSAALEKINIGAGASNIDSNIYAGSTVLTDQTSLETLNNDVVVNSNIITTAGNSGKLIF 768  
 658 -----TYNISVKTTL-----NSNGTSGNNKFSVNSNAHDN---NSLVTA-----KD 695  
 769 TNGGIGTNGIGANGAALQEVWFNCTNIGCTPANSQNETVAHSAANVVTGLTTGALKYKD 828  
 696 LA-----DYLKVKNE---TADSLPSEKVO-NG-----DNSNNAITVCKO 731  
 829 TGTIIAHGVLGDIDENNRKAGKFIKLGAMIDGVLGCVAGTLDLFDGCVNTONIGAD 888  
 732 TNGKTFNFKLKGENVNITNATGVTFVFGIDQSG-----LTTPKLTVGSDTNGN 783  
 889 -NANSISTINOIGDNTKNTVI---ANDIFVDNIHTNGGILQGLNLTTHNIDFGA---NGG 943  
 784 RLVTQEPSADCSNKNILKGLSPTLPSTASPSGRNIALGNTIEEKXNAASIDDLVLA 843  
 944 TL-----EFGNNTYLN-----NALIVQNCQILNAFTNLKASODTTGTVKIINI 988  
 844 G-----FNKNGKDKDFVSTYD-TVDFIDGNA-----TTATVTV-DEANQTSKYA 887  
 989 GOIGTPPOFTTOVNNKNLTLVSSVNSINFGDANSQILLSAPVDQTIKFNNLNETGGGI 1048  
 888 DVNVNDEKTIETLDNGKQKOLGVKTIKLTETVSTNGNAT-TFSTVD-----D 931  
 1049 ITLDSGNNLTJSGNGIK-LGSKGNELSLNIGKGVTVTVNDLIDQIHOLNTNNGALFD 1107  
 932 DHALVKASDIAGNLTLAEETHHTKGTANTALOTFTVKVDENDKADDTNAITVKGDDTS 991  
 1108 DOSLTSKAK---IKNIN-----TGVVAGGA-----TYLDAINDN---FDLN-----TS 1144  
 992 GKV-----NTLKLKGNGLDIKTDKGVTVTGI---NTQSG-----LKAGDSTLLNNGLS 1039  
 1145 GWVFKHQSILELKNSSNTN---DHTITLSALDPGNNOFGIILKLTDTNKLFDNNG-- 1199  
 1040 IKNTA-----SNEQIQVGADGVKFAWNVGV-----VGAGIDGTTTRIRD-EIGFTGTNGS 1089  
 1200 --NVAVTLGTANIMLK-----QLTFASIDNGALAKVGINVENVTNLKIDIELNEVNANVL 1253  
 1090 LDKSKPHLSKDGINA-----GKKKITNIO-----SGETAKNSHDAVTC-----GKIYD 1132  
 1254 FNKNTTVAATGNIHGVDFQGNAGVINLNDIDIEIDGVSITGTCN---VNGTLNFGSGKVTG 1311  
 1133 LKTELENKISSTAKTAQNSLHIEFSVADEQG---NNFT-VSNPYSSYDTSKT-----1179  
 1312 LINNIVMLQAGADGVLSASGNSIITEIOGNGNNLTFAANSHLTDTIDNKTGGQDLNLVLF 1371  
 1180 -----SDVTTFAG-----ENGIITTKVKNVGVVRVIGID-QT 1207  
 1372 INGSVSGSIGANAAGVDIIINAGSVNFSNLTLSGNIIVSDGATGMOVNNVTATDISGKN 1431  
 1208 KGLTTPKL-----TVGNNGKIVINSQNGQNTIT-----1237  
 1432 ANNGTLKLNHTPPNITVSTLGNNAIG-TIEVANNDVTITGLTQAQNIHFSNATQAATLT 1490  
 1238 -GLSNTLANVTNDKGSVTE-----OGN---IIDEKDKTRAASIVDVLVSAGFNLQNG-- 1287

1491 LGAASQVNTTITAGNIHTLEVTFDFTGNDGIIGDAN-NRLKSI-----ELTGNQTV 1541  
 1288 --EAVDFSVSYDVTNFGANGN---TTTAKVTYDDTSKTSKVWYDVNVDDPTI---EYDKKL 1340  
 1542 TINSPHYVSSITTANNAQGNKUNIEGGIYDGLSKISLANVQISEDTTIRGDVYSKYL 1601  
 1341 GVKTTTLTSTGTGANKFALSNQATGDALV-----KASDIVAILNTLSGD-IOTAKGASQ 1393  
 1602 NIDAGKITNDRGNNMKNPLDIPDALIDLPLRSLSLNFYFTDIKADNLNFAADTAT 1661  
 1394 AN-NSAGVYVDAD-GNKVIYDSTDNKY---YQAKINDGTVDKTEKVAKDKLVAQAQTPDGTILA 1449  
 1662 ANFKDAAVIDAHIDHNGGILKFNEDNAWLTQETKNNANII---EIASDKFM-----LL 1708  
 1450 OMNVKS-----VINKEQVNDANKKOGINEDNAFVKLEKAAASDNKTKNAAVTVG 1498  
 1709 QKNIKAAATLIADNANLVLDDNVEVNTNLNVRDILVLDLA---NYELKYTGNTINGLTLTI- 1764  
 1499 DLNVAQAOTPLTFAGDTGTAKKLGETLTIKGSDTDTNKLTDNNIGVVAGTDGFTVKLAKD 1558  
 1765 -----ITY---FDVALQKGGHILVSQGSNVDMSDL-DNLI-----IKIKAHSD 1803  
 1559 LTNLNS-----VNAGSTKIDEXGISFVDANGAKANTPVLISANGL-----1598  
 1804 ITNITSDTKHQIVKLETGAIYTPVQTKVIIDASEEQNKFKVKKWADANGLVLLTDTGGRD 1863  
 1599 DLGGKVIS-----NVGKGTQDT-DAANV---QOLNEVRNLLGLCNDNADGNQVNIADIK-- 1649  
 1864 DTGGRDTRGRGTNGCRDNCVGNISNNSNEA-----GGSSDDKN-YGITDVPVIF 1916  
 1650 DPN-----SGSSS---NRTVIRAGTVLGGKGNNDTEKATGGVQGVVDKDG-- 1692  
 1917 DPSILDYTKNNYVAGSLANOLNHVKDFGNTTDACKLLNDLGFMSPNRVTTETLRLSNR 1976  
 1693 -NANG-----DLSNVWVK-----TQKD-GSKKALLA---TYNAAGQTNVYT 1728  
 1977 INVNGLNEGVLNGIEVENFLTIDIANMDFNTAKEIGHRLBELSDANTVNGLNTNTLL 2036  
 1729 NNPRAEADIRNEQGIREFHVNDGNOEPVQGRNGIDSSASGKHSVAIGFOAKADGAAVA 1788  
 2037 NKK-----INLKRNTNQAIIAAGDE--DNIVTGIWMSFYGKIKONSKNS-A 2082  
 1789 IGROTAQGNOSIALGDNAQTGDSIAIGTGNVAG---KHSGAIGDPSTVKADNSISYVG 1845  
 2083 SGYOSNTGGGIIIGFDYNI---DNSIVIGAAVWADSKVKHKNKDCDRUKAKSNIYSIY 2138  
 1846 -----NNQFTDA-----TOTDFVGNNITVITESVALGSNAIS 1882  
 2139 GLYNLNTNFFVEAIGVYGRNKKIKYKRIITITDQIAIGKFINTFYSYELLGYNLIS 2198  
 1883 AGTHAGT-----QAQSDGTAGTTTAGATCTVKGFAG--QTAVGAVSVGASGAER 1931  
 2199 ---HRTTTPMFGMYATFKNNGYKENNTTFQNLISIKKNYVDKFETILGLNSV-----2248  
 1932 RIONVAAGEVSATSDAVNGSQL-----YKATQSIANTATNELDHRIHONENKANAGISSA 1986  
 2249 -----THYLSQDIIILPELHWFVINYQCKNKLPNIDARLD-----GIDEP 2287  
 1987 MAMASMPQAIYIPGRSMVTGIAITHN 2011  
 2288 LTTIRFKPAKITVN---LGGGISTKN 2310

RESULT 7

F81045  
 hemagglutinin/hemolysin-related protein NMB1768 [imported] - Neisseria meningitidis (C:Species: Neisseria meningitidis  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 C:Accession: F81045  
 R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. F.I., H.; Qib, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
A:Reference number: AB1000; MUID:20175755  
A:Accession: F81045  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2514 <TET>  
A:Cross-references: GB:AE002526; GB:AE002098; NID:g7227015; PIDN:AAF42109.1; PID:g722702  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB1768

Query Match 5.3%; Score 553; DB 2; Length 2514;  
Best Local Similarity 20.0%; Pred. No. 1.9e-13;  
Matches 503; Conservative 332; Mismatches 895; Indels 782; Gaps 121;

QY 4 IYKVFENKATGTFMAVEAKSHSTGGSCATQGVGS--VRTLSFARIAALAVLVIGATL 61  
DB 5 LYRVIFENKRGAVVAETTKRE---GKSCADSDSGSAHVKSVPFG----- 47  
QY 62 NGSAYAAQQTITKTEIGQTKINNTLAKGDA--LATGEASTAFG-----SLSKAQGSQAI 112  
DB 48 -----TTHAPVCRSNIFESLGLFSLCLAVGTANTAFADGIIADKAAKPTQATIL 98  
QY 113 AIGSVKPDN-----NGSNGNVGSHAKGNESIAIGGDV-----LAEG 149  
DB 99 QTGNGIPQVNIQPTTSAGVSVNOYAFDGVNKGAILNNSRSNTQTQLGGVIOGNPWLARG 158  
QY 150 DASIAIGSDDLIYLPKNDLKNFEHLKHGHEILKKTQTDGDKIKYRTPAAGHASTAVG 209  
DB 159 EAAVWV-----NQINSHSOMNGVIEVGGRAEVVIANPAG 195  
QY 210 AMSYAOCHFSNAGFTYATAEAYSLAVLAAQATKQSSIAV---GSKANAKANAF----- 259  
DB 196 IAVNGGFGFNASRATLTGPOQVQ-AGDLSGFKIROGVNVIAIGHGLDARDTDFRILSYH 254  
QY 260 -----AATATGGNTVNVNLGRVALGFGSQILDRDNT-DASAVYPL-----GK- 301  
DB 255 SKIDAPVWGQDVRVAVAGQDVVATGNAHSPILNNAANTSNNTANNCTHIPLFAIDTGKL 314  
QY 302 -----TLADQYKATRGDSTDIPIGSGNSNNNSIRKIIIVG---AGSRDITD 346  
DB 315 GMYANKITLIATAEQAGTRNOG---OLFA--SSGNAVIDANGRLVNSGTMAAANAKDITD 369  
QY 347 AVNVAQLKLVEELANRKITFKGQGDNNNSVERGLGNTLTIKGDAQTNALTEANIGVVTD 406  
DB 370 -----NTAEHKVNIRSGOVENS----- 387  
QY 407 GNLKVKVLAKELTGLTSVATNKITVSNNTNNNAELQSGGLTFSPITGKTDTKYVSIDG 466  
DB 388 -----TAVSQOQTQIHSQSIQNTGTLSSGEILIHNSGSLKNETSGTIEA 432  
QY 467 LKFTDNSNSTATKG---TTRITKKKIGFAGTND-----GVDESKPYLDNEKLVGNSTLN 518  
DB 433 ARLAIDTDLNNGKLSQTSQSKLHDAQCKMDNRGMGLQDAPTASNGS-----SNQTN 489  
QY 519 SGSLTVNNTTGNKIQVANGIKFATVANNVANTSATVGTARITEEKIGFAGTNDGVDEQ 578  
DB 490 SYNASEHSSTTTTATG-TGTATVSIISNITAPTAD-GTIR-----THGALDNS 537  
QY 579 APYLDKERLKVGRVEITDSGI-NAGNHKITGLT-----NGIANTDAVITKOLKDA 628  
DB 538 GSIIAN-----GQTDVSAQOGLNAGQIDHQLNAKGSAPDNHNGTLISDAVHIQ----- 587  
QY 629 KPTLNAGDGISINSGDLVDSSGNTITPTTYNISVKTTL----- 668  
DB 588 -----AG--SLANQN-----GNITT-RQOLEIETDQLDNAGKLLSAEIALAVSGS 631  
QY 669 --NSNCTSGNKNKFSVNAHD-----NNSLVTAKDLA---DYLKNVNETADSAL 711  
DB 632 LNNQNGEIAATNQOLI--IHDGQOSTAVIDNTNGTIQSGRQVIAQAKSLSNNGTLAANDKL 689

QY 712 -----PSFKVQNGDGNNAITVQKDTNGKTFNTLKLKNGVGNITTTNRTATGTVTFGIDQS 766  
DB 690 DIALQDDFYVERNIVAGNELSL--STRGSLKNSHTLQAGKRIRIKANNLNDNAAGNI-QS 746  
QY 767 NGLTTPKLTIVGSDTN-GNRLVIEQVPSADGNSK-----NIIKGLSPILPSTIASPSGR 818  
DB 747 GGT-----DIGTQHNLTNRGL-----DGOQTKIAGQOMNNGTG-----RIYD 787  
QY 819 NIALGNT-IEBKDSN---AASIDDDVLNAGFLNKNNGKDKDFVSTYDVFIDGNATTAT 874  
DB 788 NIAIAATRLDNQDENGTAIAARENLNLGIGQLNRENSLIYS-----GNDMAVG 838  
QY 875 VTYDEANQTSKVAYDVNVDEKTIETLTDNGKQKQGVKTI-----KITETSTNGAT 925  
DB 839 GALTDTNGQATGKAQRIHNAGATIEAA---GKMRGLGVEKLHNTNEHLTKQLVETGRE-HIV 894  
QY 926 TETSTDDHVALVKAS---DIAGNL-NTLAEIHTTKGTANTALQTFVTKYKVDENDKADDT- 980  
DB 895 DYEAFGRHELREGTOHELGSVYNDESHLRTPDGAHENWHIKYDEYKVTQKTQVTTQTA 954  
QY 981 -----NALTV-GKD--GTSKGVNTLKLKNGKGLDITKDKG---TVTVFG---INTQSG 1024  
DB 955 PAKIISGNDLTIDGKEVFNTDSOI-----IAGGN-LIVQTEKDLGHLNEQTFGEKKVPSENG 1009  
QY 1025 L-----KAGDST-----TL-----N 1034  
DB 1010 KLHSYWRKHKGRDSTGCHSEQNVTLPPEITRNISLGSFAYESHKALSHHAPSQGTLPQ 1069  
QY 1035 NGLSIKNTASNOIQVADGVKFAVMYNGVVGAGIDGTTTIT--RDEIGFTGNGSLDK 1092  
DB 1070 SNISLPTYSNTPTPSSSLYIINPVNKYL---VETDPFANYRQWLGSYDMLDSLKL 1126  
QY 1093 SKPHLSK---DGINAGKKITNIQSGEIAKNSHDAVTGGKIYDLKTELEKNKISSTAKTAQ 1149  
DB 1127 DPNNLHKRLGDYIE-QRLNEQIAEL--TGHRLDG---YONDEQFALMDNGATAA 1179  
QY 1150 NSLHEFSVADEQGNNTFVSNPYSDYTSK-TSDVITFAGENGITTKVNGVVRVIGIDOTK 1208  
DB 1180 RSM-----NLSVGIALSAEQVAQLTSDIVMLVQKE-----VKLPDGGTQ 1218  
QY 1209 GLATPKLTV---GNNGKGVINSQNGNTITG-----LSNTLA 1244  
DB 1219 TVLPVQVYVRKNGIDGKCALLSGSNTQINVSGLSNKGSIAGRNALIINTDFLDNIGG 1278  
QY 1245 NVFTNDKGSVFTBQ---GNIIKDEKTRAASIVDLSAGFNQNGEAVDFVSTYDTVN 1300  
DB 1279 RIHAQKSAVTATQDINNIGMLSAEQTL-----LLNAGNNI----- 1314  
QY 1301 FANGNTTAKVYDDTSKTSKVYVDVNVDDTTIEVKDKLGVKTTTLTSTGTGANKFALS 1360  
DB 1315 --NSQSTTASSQNTQSGST-----YLDRMAGIYI-----TGKEKGVLA 1350  
QY 1361 NQATGDALVKASDIVAHLNLTLSGDIQTAKGASQANNAG-----YYDAD 1404  
DB 1351 AQAGK-----INIIAGIISNOSEGOQTRLOAGRDINDLTVQTSKHOATHFAD 1399  
QY 1405 GNKVIYDSTD--NKYQAKNDGTVDKTEKAKDKLVAAQATPDGTLA-----QMNKSVI 1457  
DB 1400 -NHVIRGSTNEVGSSIQTKGDVILLSGNLL--NAKAAEVSSANGTAVSAKNDINISAGI 1456  
QY 1458 NKEQVNDANKQGINEDNAFVKGLEKAAASDNKTKNAAVTVG-----DLNVAQOTPL 1508  
DB 1457 NTHVDDASKHTGRSGGKGLVITDKAQSHHETAQSSFFEGKQVVLQAGNDANILGSNV 1516  
QY 1509 TFAGD-----TGTTAK-----LGETLTIKGGOTDNKL 1537  
DB 1517 SDNGTQIAGNHVRIGTFTQSOSETYHQTKSGLMSAGIGFTIGSKTNTQENQSOSEH 1576  
QY 1538 TDNNIGVAGTGDFTVKLAKDLNLNLSVNA---GGTKIDKESGFVDANGAOKANTP---1591  
DB 1577 TGSTVGLAKGDT--TIVAGKHYEQIGTSVSPGNNNTIYAQSIDIOAHKLNLSNTQTY 1634





Thu Sep 13 14:19:27 2001

Db 291 NVTFAGNLTVDVTRVNRNTPKPIEDTNRKONIALISGESAGSMYGRNKF I----- 345  
 Qy 362 RKITFKGDDNSNSV--ERGLGNTLTIKGDAQTNALTEANIGVVDGNGLKVKLAK-- 416  
 Db 346 --VTDKGAGVHQGVFAEDDI-NILTDGNSRLNKYIADYVRV-----GKDIELANGQ 398  
 Qy 417 -----ELTGLTSVATNKIIVS--NTNNAELQSGGLTFSPITGK-- 456  
 Db 399 IHADQOILNATGHVKLNDGSSVTSNNMLGISALNLEFNATYSANMLSFRVNTDKLNN 458  
 Qy 457 -----TDKTVYISDGLKFT--NDSNSIATKGTTRITKKGIGAGTNDGV 498  
 Db 459 LSKVSARAADLOSGLNLDKA--SVLAHKLTLNLSNDVSLNMQSKSANKLKIKVRD-- 514  
 Qy 499 DESKPYLONEKLVKGNSTLN-----SGSLTV--NNTTKGQIOVGANGIK 541  
 Db 515 ----LNLNNSLSANLTLNNTNLTNKKSKFTAGNLTNTVNTNVLNDSLSAANLTL 570  
 Qy 542 FATVANNVANTSATVCTARITEEKICFAGTNDGVDEQAPYLDKERL-----KVGRVEITDTS 598  
 Db 571 L-NVTKNVLNDA-----SKLSANKLDLNDVNTLNSKSTLSAGELTFPKVKNYVTLNDS 625  
 Qy 599 GINAGNHIKTGLTNGIANTDAVITQLDKAP--TLN-----AGDGISINS-----NNGD 646  
 Db 626 ELAANLSLNASHNVLNKKLSAQADIKAVNLTLNDTTELTAKNLDINSITITNNGT 685  
 Qy 647 LVDSSGNITPTVYNISVKTTLKNSNGTSGNNKFSVSNAH--DNNSLVTAOKDLADYLKYNVE 705  
 Db 686 IAGIFANITTEKLNKEKALIL--AEQNLNFTVNGSHYENKGDIVSKDKATV--TFSK 739  
 Qy 706 TADSAFPSFKVQNGDN-----SNNAITVGKDTNGKTFNTLKLKENGVNITTNRATGTV 759  
 Db 740 NSDFTSGSLVNAQNLKVNVNFTISQGD-----ITLIG-----NVTLN-ASGTF 786  
 Qy 760 TFGIDQSGNLTTPK-LTVGSDTN-----GNRLVIEQVPSADGNSTKNIKGLSPTLPSTIAS 814  
 Db 787 T-----NSGNLTVTKLDVGDIONFTKNGLTVGEDLHIKSKITN----- 828  
 Qy 815 PSGRNIALGNITIEBKSKNSAASDDVLNAGFNL-----KNN--GKDKDFVSTYD V 863  
 Db 829 -DGKLSIKNL-----NISSEADFNNGLTLGIEALKATKGNFTNKEKAILASNSLL 880  
 Qy 864 DF--IDGNATTATVYDEA--NQTSKVAYDVNVDEKTIETLDGDNKKOLGOLGVIKLT-E 917  
 Db 881 DISVAEGKKTENNCTIESGKNLNTNTGAF-LNVYDNATI-----RSGVLNITSTGN 931  
 Qy 918 TSTNG-----NATFES--TDDHALVKASDIAGNLNLTABEIHITKTGTANTALQTET 967  
 Db 932 VSNNGTLISNERLNTISAANFTNESNGTVMSN---GLLNTIAKOGNITNKNLNASRQOLN 988  
 Qy 968 VKKVDENDKADD--TNALTV-----GKD-----GTSKV-----NTLKL 999  
 Db 989 LTAADNTNDSNISNKTAVLHSLGNLSLNSKQOVNLYGELIYAGNNISVRAHOLKNDVKL 1048  
 Qy 1000 KGKNGLDKTK-DKDGTVT-----FGINTQSLKAGDSTTLNNGGISIKNTASNE 1047  
 Db 1049 MG-----DITTKKEGQASYKLYQASNGHFGNDGSSGSEGD-----LNIG----- 1091  
 Qy 1048 QIQVGADGVKFAWNVNNGVAGIDGTTTRITRDEIG--FTGTNGSLDKSKPHLSKDGGINAG 1105  
 Db 1092 -----KPADLN-----KLTQORIGKIYAGROLTFNKS-----NAG 1122  
 Qy 1106 GK-----KIT-----NIQGEIAKNSHDVATGKGIYDLAKTELENKISS 1143  
 Db 1123 GKSELLNRTINVKNLKLSYSDSVFENNMSQKV-----DLVT--KIFEAKSDIE----L 1171  
 Qy 1144 TAKTAONSLHEFSADEOGNNTTVSNPYSSYDTSKTSVDITPAGE--NGITTKVNKGV 1200  
 Db 1172 TFKT--NGHPYVL-----NFKSNNEKKYRSENKTFKXSIGDLINEALSDSAPEAIE 1223  
 Qy 1201 RVGIDOTKGLTTP--KLTVGNNNGKGVINSQNGTITGLSNTLAN----- 1245  
 Db 1224 AYSGSSSNYPINVSYLAALGNAN-----NSSNPHYLNTALKHILGNWQODLKKOENI 1277

Qy 1246 -----VTNDKGSVRTTEOGNLIKDEKTRAASIVDVLISAGFNLQNGCEAVDFVSTY 1296  
 Db 1278 KVLKOKWEDFKDKGASKMLD---LYPNTDEKAKIAGIIRNGNDTISDVESEDPKKKY 1334  
 Qy 1297 DTVNFANG--NTTTAKVYDDTSTKSVVYDVNVDDTTIEVKDKKLGKVTTLTSTGTG 1353  
 Db 1335 S--KFQNGEAKNDGTSDSYSTRASEKYKKVENVDHKE-NIDEHKLNIGKHEITVPGVS 1391  
 Qy 1354 ANKFALSNOATGDALVKASDIVAHL--NTLSGDIQTAKGASO----- 1393  
 Db 1392 FEN--LNNKNMHDQPKLGEIDKSIISELLAQPVYTEASAARDSPPRVNQNDKALDNL 1449  
 Qy 1394 -----ANNSAG-----YVDADGNKV----- 1408  
 Db 1450 RFLSYINONNVLGAKYFFENQDLTEDDKLGKIKRIGDNYFEHQLITRLIEKVADNHLTLK 1509  
 Qy 1409 --IYDST-----DNKYQAKN-----DCTVDKTKYEVAKDKLVAQAOT-----PD 1445  
 Db 1510 HGLHDIALVKKLIDSASIOAKDLNLKVGELTKEOKDLNKEDIVVYVKTVEYNGQEVLPQ 1569  
 Qy 1446 GTLAQMNKSVINKEOVNDANKKOGINE--DN-----AFVKGLEKAAASNDKTKNA 1494  
 Db 1570 VYLAKQTIIEVEKQRGVGTGQIRAGIIDVKVDVDRNTGTIAGYAVGLE---AKNKLKN-- 1624  
 Qy 1495 VTVGDL-----NAVAQTPLTFAQDTGTT----- 1517  
 Db 1625 --TGDILSQRLSKLVGKGLSTGVTVYDGTGATKVKRKARIKSEGHYILETDKKNVDLT 1682  
 Qy 1518 -----AKKLGFT----- 1524  
 Db 1683 ASELKNGTQIKAKDMLNDLIYETSYKYEKLFKNGGBEIGDRVTQTSQAKSVGTDAF 1742  
 Qy 1525 ----LTIKG--GQTDNKLNNIGVAGTGTGTVKLAKDLTLN-----SVN 1566  
 Db 1743 DHLHLSLEGDVNQTSNLKANRTTGVVKG--DENTKAGKDLFHRQIDTIVTSIGTVSASAS 1800  
 Qy 1567 AGTKIDEBKIGFISDANGQAKANTPVLSANGDLGGKVISNVGKTKDTPDAANVQOLNEV 1626  
 Db 1801 GGG--QSAGLSLDDQGVETYNKTATAGANAD-----VTNFMKRTRETETSLTSHRSEF 1852  
 Qy 1627 RNLIG-----LQNDNADGNQVN-IADIKDPNPGSSSSNRTVYKAGTVLGGKGNNDTEKLAT 1681  
 Db 1853 NALSCELVMGKADTGGVDINRDVEVITKPEEIAAEQKAAEEAKA-EVKENEASETAAK 1911  
 Qy 1682 GGVQGVKDKDGNANGDLNVMVKTQDKGSKKALLATYNAAGQTNVYTNPNPAEADR----- 1737  
 Db 1912 ETEE--AENDNVAEKDKTKPKFKLTD--BEIAAAPETKGEDEFAAYKAREEDRKKGF 1966  
 Qy 1738 -INEQIRFFHYNDQNPVQVQGRNGIDSSAGKHSVAIGFOAKA-----DGE 1784  
 Db 1967 TLSAEQIESTKARDEKTTYELKYGVGAEAHSAADAISNKARQIIDTQNGLKQDGT 2026  
 Qy 1785 AAVATGROTAQAGNOSTAIGDNAOATGDQSIATGTGNVAGHSGAIGDPPSTVVKADNSYSV 1844  
 Db 2027 VAL---QBASDVNLNATGDLAGASAKLKFELST---IEKSRGA-----SDGRSIL 2071  
 Qy 1845 GNNNOFTDATQDVFQVGNITVTSNSVALGCSAISAGTHAGTQAKSDGTACTTTTA 1904  
 Db 2072 GGRNLNA-ARGGDI--TLNNTVETENSLSLKA RDNVNVNSGV-TEQKDESNSQSLKVT 2127  
 Qy 1905 GATGTVKGFAGOTAVGASVGSAGBERRIQNVAAAGVSATSTDAVNGSQL-YKATQ--SI 1961  
 Db 2128 GASGGGVWAGGCSAG-VSANGVSGS---YNESNTTESTSHTNSLLRGKSLRVEAGKDFNL 2182  
 Qy 1962 ANATNELDHR IHONENKANAGISSAMAMASMPQAYIPGRSMVTGGTATHNG---OGAVAV 2018  
 Db 2183 ISSNVVDVH-LHL-DVKGDTNVVSKQDSYRKERGYNYSVSAGVGVSTAGGARPNGSVGL 2240  
 Qy 2019 GLSKLSDNGQWV 2030  
 Db 2241 GVSAAENNSKIV 2252

```

RESULT 10
D64962
probable membrane protein b1978 - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 02-Feb-2001
C:Accession: D64962
A:; Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A:; Rose, D.J.; Mau, B.; Shao, Y.
A:; Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: D64962
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2383 <BLAT>
A:Cross-references: GB:AE000289; GB:U00096; NID:g1788285; PIDN:AACT5042.1; PID:g1788288;
A:Experimental source: strain K-12, substrain MGL1655
C:Keywords: nucleotide binding; P-loop; transmembrane protein
F:54-70/Domain: transmembrane #status predicted <TMM>
F:1564-1571/Region: nucleotide-binding motif A (P-loop)

```

Query Match 5.1%; Score 526.5; DB 2; Length 2383;  
Best Local Similarity 22.0%; Pred. No. 1.8e-12;  
Matches 523; Conservative 302; Mismatches 933; Indels 615; Gaps 125;

Qy	47	ARIAALAVLIGATLNGSAYAQQITTK---IEIGQTNKIN-----NTLKGDALATG	94
Db	84	AQIAIANANTVPYTLGALESQAVERFGISVAELRKLNQFRTFARSFONVRQGDELD	142
Qy	95	EASTATGSLSKAOGS-----QAIA-----ICSVKPDPNNGSNGVGHSHAKGNESIA	143
Db	143	PAQVSEKKLTPPPGNSDNLLEQOIATSSQIGSLAEADNSE--QAANMARGWASQASG	200
Qy	144	---DVLAE--GDASIAIGSDDLYPKLNLDLKNFEHKLHG-----HEILKK	184
Db	201	AMTDWLSRFGCTARITLGVDE-----DFSLKNSQDFDLHPWYETPDNLFPESQHTLHRTDER	255
Qy	185	IQISTD--GKKYRTRTAQG-----HASTAVGAMSTAQGHF---SNAFTYATA	228
Db	256	TQNLNGLWRHPTFTWMSGINFFDHDLSRYHSRAGIGA-EYWRDYLKLSSNGYLRLTNW	314
Qy	229	EAAYSIAVLGLAAQATKQSSIAVGSNAKANAF--AATAIGNTVWN--LCGRVALCGFSQI	284
Db	315	RSAPELDNDYEARP-----ANGWDVRAESWLPAPWPHLGKGLVYEOYYGDEVAL--FDKD-	366
Qy	285	LDRDNNITDASA----YVPLG-KTLADQYKATROGSDTIFSI-----GNS-----NNNN	328
Db	367	-DROSNPHAITAGLNVTPPLMTFSAEQROGQOGENDTRFAVDFTWQPGSAMOKQLDPNE	425
Qy	329	SSIRRKIIINGAGSRQTDV----NYAQLKVEELANRKIITFGKGDNNNS--NSVERGLGN	383
Db	426	VAARRSL-----AGSR-YDLVDRNNNTVLEYRKELVRLTLDPTVTKSGEVKSLVSSIQT	480
Qy	384	TLTIKG--DAQTNALTEANIGVVPDGNGLKVKL-AKELTGLTSVSAT--NKITVSNNTNNN	439
Db	481	KYALKGYNVEATALEAAGGKVVTGKDLVTLPAYRFTSTPTDNTWPIEVTAEADVKNGL	540
Qy	440	AELQSGGLTFSPIGTGKTDKTVYSIDGLKFTNDNSNIATKGTTRITKKIG-----FA	492
Db	541	SNREQSMVVVQAAPTLSQKQSSV-SLSTQTLNADSHSTAT--LTFIAHDAAGNPVVGVLVS	597
Qy	493	GTNDGVDE--SKPYLDNEKLKVGNSNTLNGS-----LTVANTTGNKOIQVANGI---KFA	543
Db	598	TRHEGVODITLSDWKD-----GDGSYTQILITGAMSGTILTMPQLNGVDAAKAP	647
Qy	544	TVANNVA-NTSATVGTARITEEK-----IGFAGTND-GVDEQAAPYLDK-----ERL	587
Db	648	AVNIISSVSSRTHSSIKIDKDRYLSGNPIEVTVELRDENDPVEKQKQOLNNAVSI	707
Qy	588	KVGRVEITD-----SGINAGNH-----KITGLTNGTA-----NTDVAVTIKQLDKAKP	630

Db	708	KPG- ---VTTDWKETADGVYKAYTATYTKGSGLTAKLLMQNMWNEDLHTAGFIIDANPOS	AK	764
Qy	631	--TLNA-GDGISINSGDLVDSSGNI ---TPTYNISVKTTKLNSGTSGNK ---	678	
Db	765	IATLSASNNGLANEAATV- -SVNVADEGSNPINDHTVTFVAVLSGATSFPNQNTAKT	822	
Qy	679	-----FSVSNAHNNSLVTAKDLADYLNKVNKTADALSALPSFKVQNGDNSNNAITVG	729	
Db	823	DVNGLATFDLKSQKBDNTVEVTLENGVKQTLIVSEFGSDSSTAQVDLQK- -SKNEVVADG	880	
Qy	730	KDTNGKTFNTLKLGE- -NGVNITTTNRATGTVTFG- -IDQSNGLTTPKLTIVGSDTINGNR	784	
Db	881	NDSVTMTATVRAKGNLLNDVMVTFNWSAEAKLSOTEVNSHGDIAATATLT- -SLKNGDY	938	
Qy	785	LVIEQVPS-ADGNSTKNIIGLSPTLPSIASPSGRICALGNTIEE- -	829	
Db	939	RYTASVSSGSQAQQVNFIDQSTAAITLSVPSG-DITVNTAPQYMTATLQDKNGNPLK	997	
Qy	830	OKXSNAASIDDLNAGPNLNKNGK- -DKDFVSYYTDVDFIDG- -	872	
Db	998	DKEITFSPVNDVASKPSISNGKGKMTDSNGVAIASLTGLAGTHIMARLANSNVSDAQ	1057	
Qy	873	ATVYDEAN- -QTSKVAYDVN-VDEKTIETL-DGNGKKQLGVKTIKLT- -	916	
Db	1058	MTFVADKRAVVVLOQTSKAEIIGNGVDETTLTATVKDPNSHPVAGITVNFMPQDVAANF	1117	
Qy	917	ETSTGNCNATFSTDDDHALVKASDIA- -	965	
Db	1118	TLENGIATQANGEAHVTLKGGKAGTHVATLGNNTSDSQPVTFVADKSAQVVLQ- -	1176	
Qy	966	FTVKVVDENDKADDTNATVVG-KDGTSGKVNTLKL- -RGKNGLIDIKTKDGTWTFGI- -	1019	
Db	1177	--ISKDEITGNGVDSATLTATVKDQDFNEVNNLPVTFSSASSGLTTPGVSYNTNESGIAQ	1234	
Qy	1020	NTQSGLKGADST- -TLNNGLSIKKTASNEQIQGVADGVKFMWNGVGVGAGIDGTTRI	1076	
Db	1235	ATLAGVAFGEKVATASLANGASDNKTVHFIDGTAAKIIEIAPVDPISIIA- -	1289	
Qy	1077	TRDEIGFTGNSLDSKPHLSKDGKNAGGKKTNIQSGEIAKNSHDVATGGKIIDLKTE	1136	
Db	1290	-----NSGSVIT- -ATVDDNGFPPVKGVTV- -	1313	
Qy	1137	LENKISSTAKTQNSLHEFSVADEQGN-NFTVSNPYSSYDTSKTSOVIITPAGENGITT- -	1193	
Db	1314	---NFTSNAATAEMTNGGQAVTNEQKATVYTNTRSSIESGARPDVTEASLENGSSTLS	1370	
Qy	1194	---KVKGVVRVIGIDQTKG- -	1241	
Db	1371	TSINYNADASTAHLTLQALFDTVSAGETTSLEYKVDNNGV- -	1424	
Qy	1242	TLANTYNDKGSVRTT-EQGNIIKDEKTRAASIVDVLISAGFNLQGNCEAVDFVSTYDVTN	1300	
Db	1425	PSEGVTSPNAYITTHNDGNYFASFTATKAG- -VYQLTATLE- -NGDSMQQVTVY-VPN	1478	
Qy	1301	FANGNTTTAKVYDDTSTKSVKYVDYVNVDDTTEVKKLGVKTTTLTSTGTGANKFALS	1360	
Db	1479	VANAEITLA- -ASKDPVIADNN- -	1516	
Qy	1361	NOATGDALVKASDIAVLAHNLTSIGDITAKAGASQANNAGVVDADGNKVIYDSTDNKYYQA	1420	
Db	1517	NTEV- -	1546	
Qy	1421	KNDGTVDKTKEVA- -	1457	
Db	1547	KAKVTLKGTAKAGAHVTVASMTGGKSOLVNVNFITADTLTAOVNLNVTEEDNFIAN-NVGMT	1605	
Qy	1458	NREQVNDANKKQGINEDNAPFKLEKASDNKTKNAAVTVGDNLNAVAOTPLTFAGDTGTT	1517	
Db	1606	LQATVTDGNGNPLANEAVTTLTPADVYASFTLQOGGS-AITDINGKAEVTL- -	1559	
Qy	1518	AKKLGETLTIKG-QQDTNKK- -	1562	
Db	1660	SGTYPVTYSVNNYGVSDTKQVTLIADAGTAKLASLTSVFSVYTTTEGAT- -MTASVTDA	1717	

Thu Sep 13 14:19:27 2001

Qy	1563	NSVNAGGKTKIDKGIK-----VDANQAK-----ANTPV-----LSANGLDGGKVISN	1607
Db	1718	NGNPVEGKIKNFRGTSVTLSSSTSVETDRGFABILVTSTEVGLKTVSASLADKPEVISR	1777
Qy	1608	VKGTKDIDAANVQOLNEVRNLGLGNDAGNQVNIADIKKDP-----NSGSSNRT	1660
Db	1778	LINASADVNSATITVLSLEPEQVNVQADVAKAHVN--DQGNPVAHQPVTFSAEPSSOM	1835
Qy	1661	VIKAGTVLGGKNDTEKLAGVGVQVVDKGN--ANGDLSN--VWVTKQDG-----	1709
Db	1836	ILSONTV-----STNTQGVAE--VTMTPERNGSYMVKASLPNGASLEKOLEADEKLTLT	1888
Qy	1710	SKKALLATYNAAGTQNTVNNPAEADIRNEQIRF-----PHVNDGNOEPV	1756
Db	1889	ASSPLIGVYAPTGAFTLTLTSANGTP--VEQVINFVTPPEGATLSGGKVRTNSSGQAP	1947
Qy	1757	VOGRNGIDSSAGKHSVAIGFOAKADGAAVAIGKQTAGNQSTAGDQAQATGDSIAI	1816
Db	1948	V-----LTSNKGVTVTVTASF-----HNGVTIQOT-----TVKVTGNSSTA-	1984
Qy	1817	GTGNVAGKH--SGAIGDPSVTKADNSYSVGNNOFTD-----ATQTDVFGVGN--NITV	1867
Db	1985	-----HVASFIADPSTIAATN-----TDLSTLKAATVED--GSGNLEIGLTV	2023
Qy	1868	-----TESNVALGSNSAISAGTHAGTOAKKSDGTAGTTTATAGTVKGFAGOTAVGVS	1924
Db	2024	YFALKSGSATLSLTAVT-----DQNGIATTSVKGAMTGSVTVSAYTTAGGMQTV	2073
Qy	1925	G-----ASGAERIONVAGEVSATSDAVNGSOLYKATQSIANATNDELDRHONENKANA	1981
Db	2074	DITLVAGPADTOSVILKSNRSSLKGDYDTSALRLVL-----HDSGNPIKVS	2122
Qy	1982	GISSAMAMSPQAYTPG-----RSMVTG--GIAT-----HNGQGAVAVG	2019
Db	2123	GMEFVQSGTNPVYIKISAIDYSLNNGDYKATVGGEGEATLIPVLNGVHOAGLSTTIQ	2182
Qy	2020	LSKLSDN--GOWVPKING-----SADTQGHVGA	2045
Db	2183	FTRAEDKIMSGTVSVNGVYTDLPPTTFPSQGFPGA	2215

RESULT 11

S76109  
 hypothetical protein - Synecchocystis sp. (strain PCC 6803)  
 C:Species: Synecchocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
 C:Accession: S76109  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Oikawa, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis sp.  
 S.  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S76109  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-3029 <KAN>  
 A:Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BAAL0087.1; PID:d101073  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 5.1%; Score 526; DB 2; Length 3029;  
 Best Local similarity 20.4%; Pred. No. 2.6e-12;  
 Matches 493; Conservative 269; Mismatches 823; Indels 828; Gaps 111;

Qy 60 TLNGSAYAOQITTKIEIGQTKNNITLKG----- 88  
 Db 697 TLDANITADDIINIAESGQAIPITGVGGEFNVGDTVLTIVNDKTTFTGAVGAGGLFSINV 756  
 Qy 89 -----DALATGEASI-----AFGSLSKAAGSOATAIGSVKPPD-----NNGSNGNVGSHA 133



QY 1066 VGAGIDGTRITRD-----EIGFTGTNGS-----LDKSKPHLS-----KGINA- 1104  
Db 1649 GEMTLDYGTLLADGDEYELSVATNPTGNSATATQIVVDTTAPTPTVINAIAVDDIINAV 1708  
QY 1105 -GKKI-----TNISGELAK-----NSHDVATGKIYDLKTELEN--KISSTAKTAQN 1150  
Db 1709 EAGSPVAVSGTTGVEDGQVVVTVIDGNTYATVTCNAWTFNIPVADIANFEAEVAVTSD 1767  
QY 1151 SLHEFSVADEQGN-----NFTVSN--PYSSYDTSKTSVITF--AGE-----NGITTKV 1195  
Db 1768 -----TVSDLAGNPATPATRNITVDTTAPTPTVINAIAVDDIINAVAGSPVAVSGTTTGV 1822  
QY 1196 NKG-VYRVGIDQKGLTTPKLVGNNGKGIIVNSONGQNTITGLSN-----TLANVTN 1248  
Db 1823 EDGQVVTVTID-----GNTYATVTCNAWTFNIPVADIANFEAEVAVTSD 1870  
QY 1249 DKGSVRTTEQGNLIKDEKTRAASIVDVLSAGFNLOGNEAVDFVSTYDVFANFANGNT-- 1306  
Db 1871 LAGNPATPATRNIT--TVDTVAPAVTIDSISDDTGAQAN-----DFITNDDTLVF-NGTAEA 1923  
QY 1307 -TTAKTYIDDT--SKTSKVYDVNVDTTIEVKDKKLGKVTTLTSTGTGANKFALSQ 1362  
Db 1924 DSTVVVSLDGEIGTGTANCAGEMTLDYCTLLADGDYELSVTATNPTGNSAT----- 1976  
QY 1363 ATGDALVKASDIVAHLNTLSGD--IQAKGASQANNSAGYVDADGNKVIYDSTDNKYQA 1420  
Db 1977 ATQTIIVVDTTAPTPTVINAIAVDDIINAVAGSPVAVSGTTGVEDGQVVVTVTIDGNTYTA 2036  
QY 1421 KNDGTV-----DKTKEV-----AKDKLVAQAQTPDGTLAQMN 1453  
Db 2037 TVTGNWTFNIPVADIANFEAEVAVTSDVLADGNPATPATRNITVDTTAPTPTVINAIAV 2096  
QY 1454 KSVINKEQVN-----DANKKQGINEDNAFVGLEKA--ASDNK 1489  
Db 2097 DDIINAVAGSPVAVSGTTGVEDGQVVTVTIDGNTYATVTCNAWTFNIPVADIANFEA 2156  
QY 1490 TKNAAVTVGDLNVAQPTLPFAGDTGTTAKKLGTTITKGGQDTNK-----LTDNNGIV 1544  
Db 2157 TEEVAVTVSLDAGNPATPAT--RNITVDTVAPAVTIDSISDDTGAQANDFITNDDTLV 2212  
QY 1545 VAGT--DGFTVKLAKDLTNLNSVNAAGT--KIDEGKISFVDANGQ----- 1585  
Db 2213 FNGTAEADSTVVVSLDGEIGTGTANCAGEMTLDYCTLLADGDYELSVTATNPTGNSAT 2272  
QY 1586 -----AKANTPVLSA-----NGLDLGGKV--ISNVGKGTGD----- 1614  
Db 2273 ATQTIIVVDTTAPTPTVINAIAVDDIINAVAGSPVAVSGTTGVEDGQVVTVTIDGNTYTA 2332  
QY 1615 -----TDAANVOOLNEVRNLLGLGNDNADGNQVNTADIKKDPNSGSSSNRTV 1661  
Db 2333 TVTGNWTFNIPVADIANFEAEV-----VATVSDLAGNPATPATRNITV 2378  
QY 1662 -----IKAGTVLGGKGNNDTEKLTATGGVQVG-----VDKDG- 1693  
Db 2379 DTTAPTPTVINAIAVDDIINAVAGSPVAVSG-----TTTGVEDGQVVTVTIDGNTYTA 2431  
QY 1694 -----ANGDLSNVWVKTKQDGSKALLATYNAGQTNVYTN--PAEA 1734  
Db 2432 TVTGNWTFNIPVADIANFEAEVAVTSD-----LAGNPATPATRNITVDTVAPAVT 2485  
QY 1735 IDRINEQ--GIRFFHVND-----GNQE--PVQGRNGID-----SSAGSKHSVAIGF 1777  
Db 2486 IDSISDDTGAQANDFITNDDTLVFNGTAEADSTVVVSLDGEIGTGTANGAGEMTLDYTG 2545  
QY 1778 QAKADGEAAVAGRQTAQGN-----QSIALGDNAQATGQDSIAI-----GTGNVA 1823  
Db 2546 TLLADGDYELSVTATNPTGNSATATQIVVDTTAPTPTVINAIAVDDIINAVAGSPVAVS 2605  
QY 1824 GKHSGA-IGDPSTVKAD-NSYS-----VGN-----NNOFTD--ATQTDVFG-- 1860  
Db 2606 GTTGTGVEDGQVVTVTIDGNTYATVTCNAWTFNIPVADIANFEAEVAVTSDLAGNPA 2665

QY 1861 --VGNITVTYES-----NSVALGSNSAISAGTHAGTOAKKSDGTAGTGT 1903  
Db 2666 TPATRNITVDTTAPTPTVINAIAVDDIINAVAGSPVAVS-GTTTGE-----DQGVVTVTI 2720  
QY 1904 AG-ATGTVKGFAGQAVGAVSVGASGAERRIQNVAAGEVSATSDAVNGSOLYKATQSI 1961  
Db 2721 DGYTYATVTCNAWTFNIPVADIA-----NFEATEEVVATVSD-LAGNPATPATRNI 2771  
QY 1962 -----ANATNELD 1969  
Db 2772 TVDTVAPAVNELD 2784  
RESULT 12  
T31102  
filamentous hemagglutinin 1 - Haemophilus ducreyi  
C:Species: Haemophilus ducreyi  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T31102  
J. Ward, C.K.; Lumbiey, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.  
J. Bacteriol. 180, 6013-6022, 1998  
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.  
A:Reference number: Z20984; MUID:99030326  
A:Accession: T31102  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-4152 <WAR>  
A:Cross-references: EMBL:AF057695; NID:g3929017; PID:g3929018; PIDN:AAC79757.1  
C:Genetics:  
A:Gene: lspA1  
Query Match 5.0%; Score 513; DB 2; Length 4152;  
Best Local Similarity 20.4%; Pred. No. 1.2e-11;  
Matches 483; Conservative 344; Mismatches 953; Indels 590; Gaps 113;  
QY 2 NHIYKVFIPNKATGTFMAVAYAKSHSTGGSCATGQGVSVRTLSFARIAALAVLVIGATL 61  
Db 3 NKRYKLIFSKVRNCLVPVAENIKSAGNSGSKIAEDQEEPPSLA----- 51  
QY 62 NGSAYAQOITTKIEIGOTNKN-NLTKGDALATGEASIAFGSLSKAGGSOAI- GSVKP 119  
Db 52 ---CSLSPSSSIHGLHNSPLKVPKGLSVLLSLMPATPLLAQONTAEALNGKVV 108  
QY 120 DPNNGSNGVNGSHAKGNESIAIGDVLAEGDASTAIGDLDYLPKNL-----DKNEF 172  
Db 109 DSQHSSTRIY--EQKNDNSKDGIVVVEIANPEVDGVDNRKEFNIPNSAVENSRTE 166  
QY 173 HKLIHGHEILKKIQTSTDGKIKYRTRAQHASTAVGAMSYAOGHFSNAGTGTATATAAY 232  
Db 167 TSQLVG-KLHANIQLOKEAKLIINQVTDG-HESNIOGALEVA-----GKKADLIIVN 216  
QY 233 SLAVGLAAQATKOSSIAVGSNA-----KANAFATAIGGNTVNLGRGVALGFGSQILDR 287  
Db 217 PNGITLNGVKTINTDFRVVSTSDIIPHRENGLLSVR-NGKVTIDKG-GVATNGLSHFVV 274  
QY 288 DNNDTASAYVPLGKTLADQYKATRGDSTDFISIGNSNNSNRRKIIINVGAG-SRDT 346  
Db 275 ARNIDQKGKTVAKT--ENQKSYNPANIT--FAAGSLNLYNLTREATPISGTSRTSDTP 330  
QY 347 AVVV-----AQLKVLBELANRKTFFKGDGDNNSNSVERGLGNTLTIKGDAQTNALTE 398  
Db 331 AISADSAGSMYGNIKFV-----VTDKAGVKKHGII-----FSE 365  
QY 399 ANIGVVTGNGLVK-----LAK--ELTGLTSVSATNKTIVSNT----- 435  
Db 366 NDINIKMDGNGNASLKELYAKKDIDLAKDLELTKGQLOQANNKLIILNSTKINLRNASEV 425  
QY 436 ---NNNNAELSGGILTPITGTTKTDKTVYSDGLKFTNDSNSTATGTTTRITKKIGFAG 493  
Db 426 SADNVVVKSENALENASMSANSLDIVTKIE-----VNRSSKYSA-GTANIKASNI--- 476  
QY 494 TNDGVDKESKPYLDNEKLUKGNSTLNSGSLVNVNTTGNKIQOVGANGLIKFPATVANNVANTS 553

Db 477 -----TLDGSSVANKITLN-----VTNATLNQSKLSAKDMEL-----NVTHNI 517  
 Qy 554 ATVGTPARITEEKIGFAGTNDGVDQAPYLKDKERIKYGRVERIITDSDGINAG-NHKITGLTN 612  
 Db 518 TLNNTSKLSAQKANKLNTENLTNGEAS-LVAEKLDINAIDKITNNGTIAGLTANIT--TK 574  
 Qy 613 GIANTDAVITKOLDAKPTLNA-----GGISLN-----SNGDLVDSSGNITPTYN 660  
 Db 575 ALENDRNALILAHQNLPTVNGSHVYVNGKDIVSKDAIYTFSSNDSFTSGSKLVDQANN 634  
 Qy 661 ISVKTTLKNSGTSONNKFVSNAHNNLSVTAKDLADYLKVNVEPADSALPFEKVQNGD 720  
 Db 635 LIVNVN--NFNITOGSEII-----LHGN---VTLNAGNFTNSGNLTMKELNITSNIESFI 685  
 Qy 721 NSNNAITVKGDTNGTFTNLTALKGNGVNIITNRTATGTVTFGIDQSNGLTTPKLTGVSOT 780  
 Db 686 NAGN-LTTGKNLEVHSNTTVKNDGKL-----LHGN---VTLNAGNFTNSGNLTMKELNITSNIESFI 722  
 Qy 781 ----NGNRLVIEQVPSADGNSTKNIKGLSPLPSIASPS-----GRNIALGNITIEBKD 831  
 Db 723 DFTNNGTLGLLEALKTASGNFTNASNG-----SLASNKSLDIYGNFTNNGTIESVKS 776  
 Qy 832 SNAAS-----IDVVLNAGFNKNGKDKQVSYDVFIDGNATTATVYDEANOTSKVA 887  
 Db 777 LNTNNTYFINNATIKSYGVLAITSOGFTN-----DSNGT--VMSHDLNITSQ-- 824  
 Qy 888 YDYNDEKTIETLTDGNGKQKLGKVIKTIKLTETSTNGNATFTFDDHALVKA-SOI----- 941  
 Db 825 --ANIINKNL-LAGGOG-----LNLJT---AKGNITNDSNSTAIVLHSDNNLNAN 870  
 Qy 942 -----AGNLTIAELH-----TTKCTANTALOTFTVKKVDENDKADD 979  
 Db 871 NKVYNGEISYQAGNISVEAKLHNDVKLSGNTTTTKSGNATVKNISI-----GGGLHD 925  
 Qy 980 TNAITVCKDGTSGV-----NTLK--LKGK-----NGLDITDKD-----TVTFG 1018  
 Db 926 ANSLRVGELLNLNGKFAIDLQKVALRGKIYAGSNLTFFKA-KEGEKEQSTAAQKIING 984  
 Qy 1019 -INTQSLKAGDSTTLNNGLSKNTASNEQIQVAGDGKFKFAMVNGVVGAGIDGTTRIT 1077  
 Db 985 TINVKNLEYSNVDVNNRSMQ-VNLYEKIFNGDNPTLTL-KNGVTFK-DFSNRRR 1041  
 Qy 1078 RDEIGFTGKSLDKSPHLSKDCGINAGGKKITNIOSGEIAKNSHDAVYGGKI---YDL- 1133  
 Db 1042 RASNDGEGTKTFTDNVAHLIEAF-----SGYSGNDHRASDDGHVKSPIYLL 1090  
 Qy 1134 -----KTELENKISSTAKTAQNSLH--BFSVADEQGNFTVSNPVSSY----- 1174  
 Db 1091 VLAQAVNNTEGENYL---KTALQHIFGNMNDLTNTTNDTINDKNQKLKWEKFN 1146  
 Qy 1175 -DTSKTSVITFAGNGI-TTKVNGVVRVGIDQTKGLTTPKLTGVNNGK----- 1223  
 Db 1147 GENNHSINLIYPADEGEVAKIIFAGVLNNG--TNCVEDKVVQELNDKAKKEYEDKFAK 1203  
 Qy 1224 ----GIVINSQNGQNTITGLSNTIANVTNDK-GSVRTTEQGNIIKDE----- 1265  
 Db 1204 KEOGRFKSRFQNGEFDWAG--DWAKEGNEISYGSKEETEETKNGIKKEHTVNIKGHEIKVP 1260  
 Qy 1266 -----DKTRAASIV-----DVLISAGFNLOGNCEAVDFVSTY 1296  
 Db 1261 TVSFENLNNHQDQSDGIDKSIISSELLAQPIYVAKADVPDVPDPRVAQNDAKVEDGLY 1320  
 Qy 1297 DT-VNFAN-GNTTAKV-----TYDDTSKTSKVVYDVNVDDTTI-----EVKDKKLGVK 1343  
 Db 1321 RTRLSYVNNQNYLGAKYFFENQDTEDDKLGKIKRIGDNYFEHOLITRLEKVAADNHLTK 1380  
 Qy 1344 TTTLTSTGTGKANKFAL-----SNOATGDAIVK-----ASDVAHLNT--- 1380  
 Db 1381 -----HGLDIALVKLLIDSASTQADLNKLVGEALTKKEDKIDVYVYKTEVN 1432  
 Qy 1381 -----LSGDIOTAKGASQANNSAGYVDADGNKVIYDSTDNKYYQAKNDGTV 1426

Db 1433 AOEVLVPQVVLAKOTIEBEVEKQRCVGTGQIRAGIIV-----KVDDVRNTGTI 1480  
 Qy 1427 D--KTEKAVDKLVAQAQTPDGTLLAQNMAVSKVINKEQVND-----ANKKQGI 1471  
 Db 1481 AGYAVGLEAKNKL-----KNTGDILSQRSLKLVGKGLGTESTGYTVDETATKVRKARIK 1535  
 Qy 1472 NEDNAFV-----KGLKAASDNKTNAAVTVGDLNA--VAQTPLTF-----AGDTG 1515  
 Db 1536 SEGHIYLETDKDNVDLTASELKNGTQIKAKDLNLDIYETSYKYKYLKFKNGKEIG 1595  
 Qy 1516 -----TTAKKLGET-----LTIKG--GOTDTNKLTDNNIGVVGAGTGTFTYKLAKDLT 1560  
 Db 1596 DRVOTSOAKSVGTDASPDHLHLHLSLEGVQNTGSLNKRNTTGVVKG--DENTKAGKDLF 1653  
 Qy 1561 NLN-----SVNAGGTKIDEGISFVDANGQAKANTPVLNGLDGLGKVSIN 1608  
 Db 1654 HRQIDTVTSGYVSASASGGG--QSAGISLTDQGVETYNKTATAGANAD-----VTNF 1705  
 Qy 1609 GKGTKDTDAANYQQLNEVRNLLG-----LGNONADGNVN--IADIKKDPNSGSSNRTVIK 1663  
 Db 1706 MKRTRETETSLTHRNSEFNALSGELYVMKADIGGVDINROVEVIKTPPEIAAEQKAAEE 1765  
 Qy 1664 ACTVLGGKGNNDTEKLTATGSGVQVGVKDGNGANGDLSNVVKTQDKGSKKALLATYNAQO 1723  
 Db 1766 AKKA-EVKENEASETAAKETEE--AENDNVAEKDKTKPKFKLTD---EETAAAFETRGE 1819  
 Qy 1724 TNYVNNPAEADR-----INQGIREFHVDNGQEPVQVGRNGIDSSASGRHSVAIGFO 1778  
 Db 1820 DFEAAVKAREEDRRKGFSLSAEQIESTKARDEKETYTYELKVGCAEAEIAAADAIS 1879  
 Qy 1779 AKK-----DGEAAVAIGRQTOAGNOSIATIGDNAQATQOOSIAIGTGVNVAGKH 1826  
 Db 1880 NKARQIIDTONGLKQDGTVAL-----QASDVLNLTATGDLGASAKLKFELST---IERKS 1932  
 Qy 1827 SGATGDPSTYKADNSYVGNNOFTDATQDVEFGVNNITVTESNSVALGNSAISAGTH 1886  
 Db 1933 RGA-----SDGRSLTGLGRNLNA--ARGGDI--TLNNVETTENSLSLKARDNVNNSG 1981  
 Qy 1887 AGTOAKKSDGTAGTTTATAGATGTVKGFAGOTAVGASVAGSAGAEIRIQNVAAEVSATST 1946  
 Db 1982 V-TEOKDESNSQSLKVTAGASSGCVNAGCSAG-VSAGVSGS-----YNESNTESTSHTN 2035  
 Qy 1947 DAVNGSQL-YKATO--STANATNDELHRIHONENKANAGISSAMAMASMPQAVIPGKSMV 2003  
 Db 2036 SLLRGKSLRVEAGKDFNLISSNVDDVH-LHL-DVKGDTNVVSKODSYSRKERGVNYSVA 2093  
 Qy 2004 TGGTATHNG---QGAVAVGLSKLSDNGQV 2030  
 Db 2094 GVGVSATGAGPNSVGLGVSANENSKIV 2123

RESULT 13  
 H81193  
 hemagglutinin/hemolysin-related protein NMB0493 [imported] - Neisseria meningitidis (C:Species: Neisseria meningitidis  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 C:Accession: H81193  
 R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. Science 287, 1809-1815, 2000  
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: A81000; MUID:20175755  
 A:Accession: H81193  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2703 <TET>  
 A:Cross-references: GB:AE002405; GB:AE002098; NID:g7225708; PIDN:AAF40927.1; PID:g722 A:Experimental source: serogroup B, strain MC58  
 C:Genetics:  
 A:Gene: NMB0493

Query Match 4.9%; Score 512.5; DB 2; Length 2703;  
Best Local Similarity 20.0%; Pred. No. 7.2e-12;  
Matches 531; Conservative 345; Mismatches 893; Indels 887; Gaps 134;

QY 1 MNH-IYKVLFNKATGFMVAEYAKSHSGGSCATGQVG-----SVRTLS 45  
DB 1 MNTLYKVFVFNKRNCHMIAEYAKRE--GKNTADTQAVGILPNDIAGPAGFIHSIVLS 58  
QY 46 FARI-----AALAVLVIGATLNG-----SAYAOQIITKIEIG-----QTNKINNTLKGDLA 92  
DB 59 FSLISLLGSALILITSSSAQAQIVADKAPAAQOQPTILQTGNGIPQVNIQTPTSAG-----114  
QY 93 TGEASTAFGLSKAGSQAIATGSKVPDNGSNGVNGVSHGKNSIATGGV-----IA 147  
DB 115 -----VSVNOYAO-----DVGNRGAILNNSRSNTQTQLGWIQGNPWLA 154  
QY 148 EGDASTAGSDDLPLPKNLDAKNEPHKHLHGHEILKKIOTSTDGKTKYRRTAOGHASTA 207  
DB 155 RGEARVV-----NOINSHSSQLNGYIEVGGRAEVVIAMP 191  
QY 208 VGAMSYAOGHFNAFGYATATAEAYSLAVLAAQATKQSSIAV---GSNAKANAFATAI 264  
DB 192 AGIAYNGGFFINASRATLTTAQPOYO-AGDLSGFKIRQGVVVIAGHGLDARDTDY-----245  
QY 265 GGVTVVNLGRVALGFGSQILDRDNDNTDASAYVPLGKTLADQYKATROGSDTDIF-----319  
DB 246 -----TRILSYHKSIDAPVWGQDVRVVGQVNDVAATGDAHSPILNNA 288  
QY 320 --SIGNSNNSIRKIIINVAGSRDTPAVVNAQLKVLBELANKRITFKGDDGDNNSV 377  
DB 289 ANTSNNTANGTHIPLFALDTG-----KLGGMYAN-KITL-----ISTV 326  
QY 378 ERGLGNTLRIKDAOTNALTEANIGVVTGNGLKVKLAKELTGLTSVATNKITVSNTNN 437  
DB 327 EQA-----GIRWQGWFA-SAGNVAVNAG-----KLVN--TGM--IAATGE-----N 364  
QY 438 NNAELQSGGLTSPITGTDKTVVYSIDGLKFTNDNSNSTATGTTTRITKKIG-FAGTND 496  
DB 365 HAVSLHARVHNSGTVASODDAIHS-----QTLDNSGTVLSSG--RLTVNLRLKLNQNN 418  
QY 497 G-----VDESKPYLDNEKLKVCNPTL-----NSGSLTVNNTTGNKQIQV 535  
DB 419 GTIOARLDMSTGGLDN-----TGNITQTSQALDLVSAGKFDNSGKIGVSDVP-----QT 469  
QY 536 GANGIKFATVANNVANTSATVGTARITEBKIGFAGTNDGVDQAPYLDKRLKRVGREIT 595  
DB 470 GLN--PNPSVIPQIPSTATSGSSIVSVSK---PGSNPVSPAP---AKNYAVGRIQT 521  
QY 596 ----TDSGINAGNHKITGLTNGIANTDAVTIKOLK-----DAKPTLNAGDGISINS- 642  
DB 522 GAFDNAGSINAGQIDIAAQNGLNGSLNAAKLRVSGDSFNNVTVKGLQAH-D-LAVNTQ 580  
QY 643 ----NNGDLVSSGNTITPYN-----ISVKTWKLNSNGTSGNNKFSVNAH-----685  
DB 581 TAKNSGHLITQTKIDNRELHAGEIAANLTLIHSGRUSNDKKNIRAAHLQDLTAGLH 640  
QY 686 -----DNLSLVTAKDLAD-----YLNKVNETAASA-----LPSFKVQNGDNN 724  
DB 641 NAGNILDAGSTVTKNLRNTCKVSVARLNTBQOTLDNTRGRIEATVNIOSQOLTNSG 700  
QY 725 AITVGKDTNGKFTNTLKLKNGENVNTTNRATGTVTFGIDQNSG---LTPPKLTVGSDTNG 782  
DB 701 HITA-----TEQLTINSRVDNQKLLSANQAQLAVSDGLYNQHGEIATNRQLSI-HDKNQ 756  
QY 783 NRLVIEQVPSADG--NSTKNIILKGLSPTLSPASPSGRNIALCNTIEBKDKSNAASIDID- 839  
DB 757 NTLALN---NADGTIOSAGNV-----SLQAKSLANNGLTLAGNKLDIALTDIF 801  
QY 840 -----VLNAGFNLNKNGKDKDFVSTYDVFIDGNATTATV 875  
DB 802 VVERDLTAGQLNLSTIKRLKNTHTLQAGHTLKLNAGN-----IDNQVTGKII 849

QY 876 TYDEANQTSKVAYDVNVNDEKTIELTCDNGKKKQKGVKTIKLTETSTN---GNATTFST--- 929  
DB 850 GGEQTDITS-----EQHVDNR--GLINSDGLTHIGAGQ-TLNTGTGKIYGNHIALDAQIL 902  
QY 930 ----DDHALVKASDITAG--NLNTLAEIHTTKGTANTALQTFV-KKYDENDKADDTNA 982  
DB 903 LNREETTESGTAGATAARKRLDIGAKEIHNQEGALLSSEGI FAVGNRLDEQIHA -----957  
QY 983 ITVGKDGTSKVNTKLKGKNGLIDKTD-----TVTFGINTQSLKA-----1010  
DB 958 -----AGMADTF-VNGSAGLEVQGDALMSVRNNQNNINNHFKTETYLAKAEKQVRDIT 1008  
QY 1011 -----KDG-----TVTFGINTQSLKA-----1027  
DB 1009 VLQNTYYOAGRDGLFDNSQGQKQDQTATFHLKNGSRIEANQHWYDIHYETIKERIEN 1068  
QY 1028 -----GDSITLNNGLSIKNTASNEOIOGA-----DGVKFAVNVGVVAG-----1069  
DB 1069 RPAHITVGGDLTASGQNWLN-----KDSRIVVGGRIITDOLNKEITNOSTTCKGRTDV 1123  
QY 1070 -----IDGTTTRTRDE-----IGFTGTNGSLDKSKPHLS 1098  
DB 1124 GTOWDSVTKKGWYSGRKQRORRTERNHPTPHDTQLTDFDTPVSVIOQNAASPSFOPAAS 1183  
QY 1099 K-----DGINAGKKITNIOSGEIAKNSHDAVT--GGKIYDLKTELENK-----1140  
DB 1184 AIKILIDGVSTAANGQRIHTGNVNLNATVTLPNSSLY--TTHPDNKGWLVDTPQFAD 1241  
QY 1141 ----ISSTAKTAQ-----NSLH-----EFSVADEQGNFTVSNPYSYDTSKTSDDVI 1183  
DB 1242 YRWLGSDYMLQOLOLDTNHLHRLKRGDGYEQKLVEQIHLTGYYRLDGY-----RSDEE 1297  
QY 1184 TFAG--ENGITTKVKNKV--VRVGIDQTKGLTT-----PKLT 1216  
DB 1298 QFKALMDNGLTAAKTGTLPGIALSAEQVARTLSDIVWMENQVTLSDGSTQTVLPVKY 1357  
QY 1217 V-----GNNNGK-----IVINSQNG-----QNTITGLSNTL--ANVTNDKGVVRTEQG 1259  
DB 1358 ALARKGDLNTSGGLISAEQVLLKLQNLNLTNSGTIAGRAVLQIARNINSNGIQADQIG 1417  
QY 1260 ----NIKDEKTRAASIVDLVSAGFNQNGEAVDFVSTYDTVNFANGNTTAKV---1311  
DB 1418 LKAEKINIDGGVQVAGRLTAQAQINLNG-----TTQTSGERNENTAIIDRMAGI 1469  
QY 1312 ----TYDDTSKTSKVY-----DVNVDDDTI--EVKDKKLGVKTTTLTSTGTGANKF 1357  
DB 1470 NVVGSHTEQVDNRTSGILSLHASDINLNAATVSNQVKD-----GTTQITAGNNLNLG 1523  
QY 1358 AL---SNQATGDALVKASDIVAHLNLTLSGDIQPAKA-SOANNS-----AGYVDADGNKVI 1409  
DB 1524 TIRTEHREAYGTLDDEHHRHROSQTEVSSIRTPQALLRAGNDLKTROGELEAEEGKV 1583  
QY 1410 Y----DSTNKYYQ-----AKNDGTVDKTK-----EVAKDKLVAQA 1441  
DB 1584 LAAGROVTISEGRQITELDTSVSGSKGILSSTKTTHDRYRFSHDEAVGNSIGGKMLVAA 1643  
QY 1442 Q-----TPDGTLAQ-----MNVKSVINKEQVND--ANKQGINEDN--AFVKGLE 1482  
DB 1644 GQDINVRGSLNISDKGIVLKAGHDIDISTAHNRYTNEYHESKKSVMGGLGFTIGNR 1703  
QY 1483 KAASDNKTNAAVT---VGDNLNAVAQTPPLTFAGDTGTTA-----KKIGETLTIKGGQDTN 1535  
DB 1704 KTDDTDRNTNIHVTGSIIGSLN-----GDTVTAGNRYRTGTSVSSPEGR---N 1750  
QY 1536 KLTDDNIGVYVAGTGGTFTVKLAKDLTLNLSNVNAGTKIDEKGISFVDANGQAKANTPVLSA 1595  
DB 1751 TVTAKSIDVEFANN-----RYATDYAHTQ-----EQKGLT-----VALNVPPVQA 1790  
QY 1596 --NGLDLGKVISNVGKTKDTPDAAN-----VOQLNEV--RNLIGLNDNADGNQVN 1643  
DB 1791 AQNFIOAQNVGKSKNKRNVNMAAANAQWQSYQATQOMQOQFAPSSSSGOGONNQSPSIS 1950  
QY 1644 IADIKKDPNSGSSSNRTVIK-ACTVILGGKGNNDTEKLTATG-GVQ-----VGVDKGNANG 1696

A:Experimental source: strain K-12, substrain MG1655	
C:Genetics:	
A:Gene: ydbA_2	
A:Start codon: GTG	
Query Match 4.8%; Score 499.5; DB 2; Length 2020;	
Best Local Similarity 20.8%; Pred. No. 1.5e-11;	
Matches 405; Conservative 224; Mismatches 663; Indels 657; Gaps 100;	
Qy 284	ILDRNNNTDASNYPLGKTLADQ-YKATROGDSITDIFSIGNNNNNSSIRKLIINVGAGS 342
Db 147	ILDKTKT-----LITRDSVFTYENADGT--ISLQDSNGRKATI----- 184
Qy 343	RTDVAVNVAQLKLVELANRKITFKGDGDNNSNSVERGLNLTIKGDAQTNALTEANIG 402
Db 185	-----NLWQI-----DEANNTVALEGSADGATKWQYNHGELVITGD---NATVNNNGK 231
Qy 403	VYTDGNGLKVKLAKELTGLTYSATNKITVSTNNNAELQSGGLTFSPITGTCTKTDKTVY 462
Db 232	TTVDG-----KDSGTG-TEIN-----GNGKVIODGDLVSG----- 261
Qy 463	SIDGLKFTNDSNSIATKGTTRIT-KKIGFAGTNGDVDESRYLDNEKLKVGNSLTNSGS 521
Db 262	GCHGIDITGDSATVDNKGTMTVTDPESMGIQ-----IDGDKAIVNE---GESTITNGG 312
Qy 522	LTVNTTGNKQIQVGANGIKFATVANNVANTSATGTARITEEKIGFAGTNGDVDEQAPY 581
Db 313	-----TGTO-----INGDD--ATANNNGT-----TVDGKDSGTGE 341
Qy 582	LDERLKVGRVETTTDSGINAGNH--KITCLTNGTANTDAVTIKQLKDAKPTLNAGD--- 636
Db 342	INGNN--GKVIODGDLVSGGGHGDITG-----DSATV---DNKGTMTVTDPE 386
Qy 637	-GTSIN-----SNGDLVDSSNITPTYNISVKTTLN-SNGTSGNNKFSVSAHNN 688
Db 387	IGIQVDGDAQVAVNNEGESAITNGG-----TGTQINGDDATANNNGKTTVDCKDS- 435
Qy 689	SLVTAADLADYLNKVNETADSLPSPKQVQNGDSNNAITVGKDTNGKTFNTLKLKGENV 748
Db 436	-----TGTEIAGNNGKV-----IQDGD-----LDVSG-----GGHGI 462
Qy 749	NITVNRAT---GTVTFGIDDSNCL-----TTPKLTVGSDTNGNRLVIEQ 789
Db 463	DITGDSATVDNKGTMTVTDPEISGIQIDGQAIVNNEGESTINGGTGTGTOINGN----- 516
Qy 790	VPSADGNSKTIKGLSPTLPSPASPSG-----RNIALGNTIEEKDK- 831
Db 517	-DATANNNGKTTVDGKDSGTGKTAGNIGIVNLGSLVTVGGAHGVENIGONGTVNNKGD 575
Qy 832	--SNAASIDDLVNA-GFNLKNGKDKDFVSTYDVFIDGNATATATVYDEANQTSKVAY 888
Db 576	VVSDTGSIGVLINGEGATVSTGDN-----VSNEATGFSITTSNGKVSLSAGSM 624
Qy 889	DYNVDEKTIELTGDNKKOLGVKTKIKLTETSTNGNATTFSTDDHDLVKAASDIAGNLT 948
Db 625	QVGFSTGVLDLNGNNNSVTLAALKLVVQKATG-----INVSGDANTVNT 671
Qy 949	ABEIHHTTGTANTALQTTVKKVDENDKADDTNATVTK-----DGTSGKVTLLKG 1001
Db 672	GNVLVDKDKTADNAAEYFDPVSVGINVYSDNNVTLDGKLTVSDSEVTSQSOLF----- 727
Qy 1002	KNGLDKTKDGTVTFTGINTOSGLKA-GDSTTLLNNG-----LSIKNT-ASNEQIOVGADG 1055
Db 728	-----DGSA-----EKTSGLVVIGDGTVNNMGLELIGEKALADGSDQVTSRTG 773
Qy 1056	VKF--AMVNWGVGAGIDGTTTRITRD-EIGFTGTNGSLDKSKPHLSKDGINAGCKITNI 1112
Db 774	YSVTSIVVVGSESSVYLVNGDITISGEFFPLGFAGV-----IRVQDKALLEI 818
Qy 1113	QSG-----EIAKNSHD-----AVTGGKIYDLKT-ELENKISSATAQNSLHEFS 1156
Db 819	GSATILTMQDIDISFEHGHGRTTPELTIVADSGAKIVNKGTVFION----- 861



Db 2308 NNOAGTFESSAGSLVTSRGEINNOGRLVTD--GATLSTGFDNSRAGLVSAKGAVALRT 2366  
 QY 761 FGIDQS-----NGLTTPKLVGSDTNGNRLVIEQVPSADGNSTKNIIG--LSPTLPSTIAS 814  
 Db 2367 GALNOKSGSIGGNTGVTIVAGLVNDRGREGRIKSTKGLDAN-----LKGLLOQGGGSLVG 2421  
 QY 815 PSGRNIAL--GNTIBEKDKSNAASIDDDVLNAGFNKN-----GKDKDFSVYDVTVD--- 864  
 Db 2422 ERGVTLDLNGGTLNHDGLVSTPGALLRQLGMVDNSVSGEISSRAFTFLAANTLNNOG 2481  
 QY 865 --FIDGNATTATV--TYDEANOTSKVAYD--VNVDEKTIELTGDNGRKKQLGVKTIKLTETS 919  
 Db 2482 CRLISSEALTURIKAKTLNLSKCOVLATDGLAIESQVL-----DNRAGTIGSK----- 2529  
 QY 920 TNGNATTFSTDDHALVKASDIAGNLTAEIHTTKGTANTALQTFTVKKVDENDKADD 979  
 Db 2530 --GDARISVTSLDNAEGSLVSEGRLELVADQV-----VNTL-----KLCKKNGLDIKTKDGTVTFTGINTOSG 1024  
 QY 980 TNAITVGKDTSGK-----VNTL-----KLCKKNGLDIKTKDGTVTFTGINTOSG 1024  
 Db 2561 -----SNGNOGRTAARGVLEAVGTLLOQGGELVSGGSLDLRAD---TLD---NSQSG 2607  
 QY 1025 LKAGDSTTLNNGLSIK-----NTASNEQIQVGADGVKFAWVNG--VVGAGIDGT 1073  
 Db 2608 LITAA-----NCGIALEARQVDNRAGEISSTSKVAVNA---REQLDNRGGKVIG---DSG 2655  
 QY 1074 TRITREIGFTGNSLDKSPHLS--KGINAGGKKTINIOSGEI--AKNSHDAVTVGKIY 1131  
 Db 2656 LRLTVQRL-----LNOAGKVLGARDGLSDGELFNGGGRLDSONLSVSLGGVLD 2707  
 QY 1132 DLKTELENKISSTAKTAQ-----NSLHEFSVADEQGNFT-----VSNPYSSY 1174  
 Db 2708 NOGGALVSEGSITARAARLDNRGGTFSSAGALATSOAVLDNQGGRLSDAGVTLKGASL 2767  
 QY 1175 DTSKTSVDITFAGENGITKV-----NKGVRV-----GIDOTKGLTTPK 1214  
 Db 2768 DNSR--SGVISAKGAVDIRTVGLDNRNGGIGSNAGITILVAARLDNGOQGRVSAKGLLDAN 2826  
 QY 1215 LTVGNNGKGVINSQNGQNTITGLSNTLAN-----VTNDRKGSVRTTEQGNII 1262  
 Db 2827 LKGLDORGGLVSEGTGVTDLNG--GTLVNRDGGIATPGALLRQLGAVDNGAGGEIS 2884  
 QY 1263 KREDKTRAASIVDVLSAGNLOGN-----CEAVDFVSTYDTVNFANGNTTTAKVTYDD 1315  
 Db 2885 SDRAFTLAAASLD--NRGRLIGADSLTLRIAQALD--NSLAGVISAAG--LDIAARLDN 2940  
 QY 1316 TSK---TSKVYVDVND---DTTIE--VKDKKLGKVTTLTSTGTG-----ANKPA 1358  
 Db 2941 SAKGTLASRAGIDRVDGALDNHAEGTVSCARLTASASLDNSGKLLSGNAGLSVATCA 3000  
 QY 1359 LSNQATGDAL-----VKASDIVAHLNTLSGDIQTAKGASQANNSAGYVDAGNKVI--- 1409  
 Db 3001 LDNAEGGQLISQGVLDVSSADLDNRGALSQKSLSAANLDNRGGLTSDGELELTAG 3060  
 QY 1410 -YDSTDNKYYQAKND-----GTVDKTKVAKDKLVAQAQTPDGTTLAQMNVKS 1456  
 Db 3061 RVDSADGGEISARGDRLTVERLVQROGRIGERGVSRLRGDDLDNOGGLISARGPLSI 3120  
 QY 1457 INKEQVNDANKQGINEDNAFAVLEKAAAS--DNKTKNAAVTVGDILNAVAQT----- 1506  
 Db 3121 ---ERLNVLDNRQG--GEIYSQOGFELLARRIDNNGOQRIISAGKRLRLDADALGNAGAGL 3175  
 QY 1507 -----PLTFAGDT-----GTTAKKLE-----TLTIKGQOTDTNKLTDN 1540  
 Db 3176 LSGWQGLTVTGGSLDNRGAGGTLSSKDGELAISLGGALDNHIGQALVSKGAQRIDAASLDN 3235  
 QY 1541 NIGVVAGTGTFTVKLAKDLTNLNSVNAAGTKIDEKGISFV-----DANGQAKANTPVL 1593  
 Db 3236 AQGIVSGESDVTLSIAGKLDN---GQGLVSAQRALESFERDDTLNNAAGRINGGSL 3291  
 QY 1594 SANGLD--LGGKVIN-----VGKGTKTDANVQQLNEVRNLLGLGNDNADGNV--I 1644  
 Db 3292 KGASLDNSDGLISQGRDLATIGGALVNAGAARLASGGDLL--LRSASVDNRGGKLVSOGL 3350

QY 1645 ADIKKDPNSGSSNRTVIKACTVL---CGKGNNDTEKL---ATGGVOVGVDKDGANGDL 1698  
 Db 3351 LEISAGSLDNRASGTLASQADMSRLGGALRNQODGLIFSQAQALEVQAGSLDNRQGLT 3410  
 QY 1699 SNVWVKTKQKSGKALLATYNAAGQTNVVTNNPAEADIRINEQOIRFEFHVNDGNQBPVQ 1758  
 Db 3411 -----QAQCDNRRLRIGGALDNOAGRLDSRAGN-----LDLOSGLD---N 3447  
 QY 1759 GRNGIDSSASGKHSVAIGFOAKADG-----EAAVAIGROTQAGNOSIATGDNQAQATGD 1811  
 Db 3448 GAGGVNLSAKGWLKLVGLFDNSAGVTOASLEIRAGOGVNRNOQHLSALGGDNRIVTAD 3507  
 QY 1812 -----QSIAGTGNV--AGK----- 1825  
 Db 3508 FDNQGGGLYASGLLSLDGGRFLNQGAAAGOGGKVGAGRIDFSLAGALANRFGQLESESEL 3567  
 QY 1826 -----HSG-----AIGDPSTVK-----ADNSYSV--GNNNQFTDATQTDVFGVGNNTIV 1867  
 Db 3568 HLRAAIDNSGSLRALGRSGSTRVLVAGDLNNAYGVLESANQDLQLGSLANAGRILH 3627  
 QY 1868 TBSNSVALGNSAISAG-----THAGTOAKKSDGTAGTTTATAGTGVKGFAGQTAAGVAVS 1923  
 Db 3628 TGNFTGLDQGOVIRAGGELTTNGLDIRASEWTNSSVLQAGRLNLDICTFOTAEGLL 3687  
 QY 1924 VASGAERRIONVAAGEVSATSTDAVNGS---OLYKATQSIANATNELDHRHONENKAN 1980  
 Db 3688 AVQSFTGR-----GGDWSNDGLLASNGSLRLSELGGYRGNGRATSLGDFAL---NAAS 3737  
 QY 1981 AGTSSAMAMASPMQAVIPGRSMVTGCIATHNGOGAVAVGLSKLSDNGQWVFKINGSADTO 2040  
 Db 3738 LDLCNAASLA-----GGANVTLGAGNLLVNRGRITAAGDLV---ASAASL 3779  
 QY 2041 GHVGAAGVAG 2050  
 Db 3780 NNYGTILGGG 3789

Search completed: September 13, 2001, 12:41:04  
 Job time: 32730 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:52:48 ; Search time 47.77 Seconds  
(without alignments)  
1472.190 Million cell updates/sec

Title: us-09-361-619-9  
Perfect score: 10356  
Sequence: 1 MNHLYKVFENKATGTFMAVA.....NGSADTQGHVGAAGAGGHHF 2053

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	608	5.9	2249	1 190K_RICRI	P15921 rickettsia
2	496	4.8	2003	1 YDBA_ECOLI	P33666 escherichia
3	456.5	4.4	1643	1 OMPB_RICPR	Q53020 r outer mem
4	455.5	4.4	1656	1 OMPB_RICJA	O08653 r outer mem
5	453.5	4.4	1577	1 HLYA_PROMI	P16466 proteus mir
6	443.5	4.3	1654	1 OMPB_RICRI	Q53047 r outer mem
7	439.5	4.2	1902	1 P2P_LACLA	P15293 lactococcus
8	436.5	4.2	1645	1 OMPB_RICTV	P96989 r outer mem
9	432.5	4.2	2334	1 WAPA_BACSU	Q07833 bacillus su
10	423.5	4.1	1902	1 P1P_LACLC	P16271 lactococcus
11	423.5	4.1	1902	1 P3P_LACLC	P15292 lactococcus
12	410.5	4.0	1848	1 CBPA_CIOCL	P38058 clostridium
13	404	3.9	1608	1 HLYA_SERMA	P15320 serratia ma
14	404	3.9	3591	1 FHAB_BORPE	P12255 bordetella
15	399	3.9	1902	1 P2P_LACPA	Q02470 lactobacill
16	373	3.6	1300	1 120K_RICRI	P14914 rickettsia
17	373	3.6	1325	1 YDEK_ECOLI	P32051 escherichia
18	366	3.5	1569	1 YPJA_ECOLI	P52143 escherichia
19	355	3.4	1286	1 AIDA_ECOLI	Q03155 escherichia
20	345	3.3	1183	1 CNA_STAAU	Q53654 staphylococ
21	339	3.3	1694	1 IGNO_HAEIN	P49669 haemophilus
22	337.5	3.3	2628	1 HAGA_FORGI	Q51845 porphyromon
23	335	3.2	1702	1 IGAA_HAEIN	P45384 haemophilus
24	334.5	3.2	1419	1 ALA1_CANAL	O13368 candida alb
25	334	3.2	1567	1 ICEN_XANCT	P18127 xanthomonas
26	334	3.2	1829	1 FRPC_NEIMC	P55127 neisseria m
27	330	3.2	1176	1 SLAP_BACSH	P38537 bacillus sp
28	330	3.2	1849	1 IGAA_HAEIN	P45386 haemophilus
29	327	3.2	1829	1 FRPC_NEIMB	Q93955 neisseria m
30	315	3.0	1025	1 SLAP_CAUCR	P35828 caulobacter
31	313	3.0	1541	1 IGAL_HAEIN	P42782 haemophilus
32	312.5	3.0	1256	1 MRP_STRSU	P32653 streptococ
33	312	3.0	1322	1 ICEA_PANAN	P20469 pantoea ana

RESULT 1

ID	190K_RICRI	STANDARD;	PRT;	2249 AA.
AC	P15921;			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-APR-1990 (Rel. 14, Last sequence update)			
DT	01-OCT-1994 (Rel. 30, Last annotation update)			
DE	190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).			
OS	Rickettsia rickettsii.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			
OC	Rickettsiaceae; Rickettsiaseae; Rickettsia.			
OX	NCBI_TaxID=783;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=R;			
RX	MEDLINE=90354033; PubMed=2117568;			
RA	Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;			
RT	"A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences.";			
RL	Infect. Immun. 58:2760-2769(1990).			
CC	-!- FUNCTION: ELICITS PROTECTIVE IMMUNITY.			
CC	-!- PTM: GLYCOSYLATED (POSSIBLE).			
CC	-!- DISEASE: PROBABLY ROCKY MOUNTAIN SPOTTED FEVER.			
CC	AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
EMBL	M31227; AAA26380.1; -			
PIR	A41477; A41477.			
Antigen; Repeat; Signal;	S-layer; Glycoprotein.			
SIGNAL	1 28			
CHAIN	29 2249			
DOMAIN	212 1180			
FT REPEAT	212 286			
FT REPEAT	287 358			
FT REPEAT	359 430			
FT REPEAT	431 505			
FT REPEAT	506 577			
FT REPEAT	578 652			
FT REPEAT	653 724			
FT REPEAT	725 799			
FT REPEAT	800 874			
FT REPEAT	875 949			
FT REPEAT	950 1021			
FT REPEAT	1022 1093			
FT REPEAT	1094 1165			
FT REPEAT	1166 1180			
FT REPEAT	2249 AA; 224333 MW; A9D6646C089DF087 CRC64;			
SEQUENCE				

ALIGNMENTS

34	311	3.0	918	1 YMTB_CAEEL	P34487 caenorhabdi
35	306.5	3.0	1592	1 GTF2_STRDO	P27470 streptococ
36	305	2.9	1258	1 ICEN_ERWHE	P16239 erwinia her
37	297.5	2.9	1039	1 AG43_ECOLI	P39180 escherichia
38	297	2.9	1545	1 IGA3_HAEIN	P45385 haemophilus
39	295	2.8	1228	1 SLAP_BACST	P35825 bacillus st
40	295	2.8	1341	1 VG37_BPT2	P07067 bacterioph
41	292	2.8	2491	1 TALA_DICDI	P54633 dictyostell
42	291.5	2.8	3562	1 PGCV_CHICK	Q09053 gallus gall
43	284.5	2.7	933	1 SLAP_CAMFE	P32927 campylobact
44	284	2.7	1537	1 FLOI_YEAST	P32768 saccharomyc
45	284	2.7	1861	1 APU_THETU	P38636 t anylopull

Query Match 5.9%; Score 608; DB 1; Length 2249;  
Best Local Similarity 22.5%; Pred. No. 1.le-14;  
Matches 491; Conservative 224; Mismatches 765; Indels 704; Gaps 114;  
QY 82 INNTLKDALATGEASTAFGSLSKAQSQAIAIGSVKPDENNGSNGVSH-----132  
DB 13 IOGLKKAALFTTAAIMLS-----SGALGVAIVGVIATNNAAFSNVGNWNNWEITAAAG 68  
QY 133 -----AKGNESIAIGDVLAEAGDASTAI-----GSDDLYLPKNDLKNFEFKLI 176  
DB 69 VANGTPAGGPQNNAFVYGGDYTTADAADRIIKAINVAGT-----TPVGLNI-----116  
QY 177 HGEHILKIKIQTGKIKYR-----TRAQGHASTAVGAMSYAQGHFSNA-----FTT 224  
DB 117 -----TONTVVGSIITKGNLLPVLNAGKSLTLNGNNAVAANHGFDAPADNYTGLGN 168  
QY 225 YATAEAYSLAVGLAAQATKOSSIAVGSN-----AKANFAATAIGNT-----VVNL 272  
DB 169 IALGANAALII-----QSAAPSKITLACNIDGGIITVKTDA-AINGTIGTNTALATVNV 223  
QY 273 GRGVALGFGSILDRDN--TDAASVPL---GKTLADQYKATROGSDTDFSI-----321  
DB 224 GAGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGLNLGALSQV 283  
QY 322 -GNSNNNSIRRKIIINVAGSRDID--AVNVAQKLVVEELANKIIT-----FKDGDGNS 374  
DB 284 TGDIGTNS---LATISVGAGTATLGGAVIKATTTKLTDAASAVKFTNPVVVTAIDNTG 340  
QY 375 NSVERGL---GNLTIKGDAOTNALTENIG---VVTGNGCLKVAKELTGLTSVSAT 427  
DB 341 NA--NNGIVFTGNTSTVGNVGNLTALAVVNGAGLLQVGGVVKANTLNTDNASAVFT 399  
QY 428 NKIVSTNNNAELQSGILTF---SPITGTRDKTVYSIDGLKFTNDSNSTATKG-----480  
DB 400 NPVVVTAIDNTGNANNGIVFTGNTSTV---DIGTNTALATVNGAGTATLGGAVI 454  
QY 481 ---TTRITKTKKIGFAGTNDGDESKPYLDNEKLKVGNSLTLSNGSLTVNNTTKNQIQV-G 536  
DB 455 KATTTKLTNAASVLTLTN-----ANAVLTG---AIDNTTGGDNVGLN 494  
QY 537 ANGKFKATVANNVANT-----SATVG-----TARITEKIGFAGTNDGVDQ 578  
DB 495 LNG-ALSOVTGNIGTNSLATISVGAGTATLGGAVIKATTTKLTDAASAVKFTNPV---550  
QY 579 APYLDKERLKYRVEIR--TDSGI--NAGNHKIT---GLTNGIAN-----616  
DB 551 -----VTGAIDNTGNANNGIVFTGNTSTVTDGIDGNTNSLATISVGAGTATLGGAVI 601  
QY 617 -----TDAVITIKOLDAKPTL-----NAGDISINNSNGDLVDSGNI--TPTTYN 660  
DB 602 KATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGLNLGALSQVTDGIDGNTNSLAT 661  
QY 661 ISVKTTLKNSGTSNKNFVSNAHDNLSLTAKDLADYLNKYNETADSALPSEKQVONGD 720  
DB 662 ISVGA---GTATLGGAVIKATTTKLTNAASVAVK---FTNPVVVVG-----AIDSTG 706  
QY 721 NSNNAI---TVGKDRNGKFTNTKLKLGENVNITNATGTVTGDQSGNCG---LTPK 773  
DB 707 NANGIVFTGNTSTVTDGIDGNTNAL-----ATVNVGAGTATLG-----GAVIKATTTK 754  
QY 774 LTVGSD-----TNGNRLVIEQVPSADGNSKTKI-----GLSPPLPSIASPSG 817  
DB 755 LTNAASVLTLTNANAVLTGAIDNTTGGDNVGLNLGALSQVTDGIDGNTNSLATISVGAG 814  
QY 818 RNIALGNTIEEDK--SNAASIDDLVNLGAFNLKNGKDKDFSVYDVTDFIDGNATTATV 875  
DB 815 TATLGGAVIKATTTKLTNAASVLTLTNANAVLTG---ADVNTTGGDNVGLNLGALSQV 871  
QY 876 TYDEANQTSKAVDVNVDEKTIETLTDGNGKKGKOLGVKTKILTE-----TSTNGNATFTSD 930  
DB 872 TGDIGTNSLATISVGAGTATL-----GAVIKATTTKLTNAASVLTLTNANAVLTGAI 925  
QY 931 D-----DHALVKASDIAGNLTALAEIHTTKGTA---NTALQTFVTKKVDEN 974

DB 926 DNTTGGDNVGLNLGALSQVTDGIDGNTNSLA--TISVGAGTATLGGAVIKATTTKLTDA 984  
QY 975 DKADDNTNAITV--GKDTSGKVNLTLLKGRGLDIKDKGTGTGINTOSGLKAGDSTTL 1033  
DB 985 SAVKFTNPVVVTAIDNTGNAN-----NGIVTF-----TCNSTVT 1019  
QY 1034 NNGLSITKNTASNEQIQVGADGVKFAWNVNNGVVGAGIDGTTTRIDRDEIGTGTN-----G 1088  
DB 1020 GNVG-----NTNALATVNVGAG---LLQVQGVVKA---NTINLTDNASAVFTNPVVVVG 1069  
QY 1089 SLDRSKPHLSKDGINAGGKKTITIQSGEIAKNSHDAVTGKIIDLKTLELENKISSTAKTA 1148  
DB 1070 AID-----NTG-----NANGIVFTTGNSTVITG-----NVGNT-----1097  
QY 1149 QNSLHESVA---DEGNFTVSNPYSSYDTSKTSVITFAGENGITTKVNGVVRVGI 1204  
DB 1098 -NALATVNVGAGLLQVQGVVKA---NTINLTDNASAVTF-----TNPVVVTAI 1143  
QY 1205 DOTKGLTTPKLTGVN--NNGKGIIVINSONGONTIING--LSNTLANVTNDKGSVRTTEOGNII 1262  
DB 1144 DNT-----GNANG-----IVTFTGNTSTVTDGIDGNTNALATVNVGAGITLQAGSL 1189  
QY 1263 KDEDKTRAASIVD-----VLSAGENLOGNEAVDFVSTYDTPNFANGNT-----1306  
DB 1190 -----AANNIDPFCARSTLEBENGPLDGGKRAIPY---YFKGAIANGNATLNVNTKLLT 1239  
QY 1307 ---TTAKVYTDTSKTSKVYDVNVDDTTI-----EYKDKKLGKVTTLTSTGTGA 1354  
DB 1240 ASHLTIGTVAEINIGAGNLTFTIDASVGDVLTILNAQINFRARDSVLLSNL-----TGCV 1295  
QY 1355 NKFALSNOATGDALVKASDIVAHLNLTLSGDIQT-----AKGASQANNSAGVY---DAD 1404  
DB 1296 NNTLL-----AADLVAP---GADECTVVGNGVGLNGLVNGVAGTARNIGDGG 1340  
QY 1405 GNR-----VIYDS---TDNKYQAKNDGTVDKKEVAK-----DKLVA 1439  
DB 1341 GNKFTLLIYNAVITITDDVNLEGIQNLINKNADFTSSTAFNAGALQINDATYTTDANNG 1400  
QY 1440 QAOTPDGTL--AQMNVKSVINKQOVNDANKKQINEDNFAVKGLEKAASDNKTKNAVTV 1497  
DB 1401 NLNIPAGNIQFADADAQLVLQNSGNDRTITLCANID-----PDNDEGIVI---1447  
QY 1498 GDLNAV--AQPLTFAGDTGTT---AKLGETLIG--GOTDTNKLTDNNIGVAGTGDGT 1552  
DB 1448 --LNSVTAGKLTITAG--GKTFGAHKL--QTILFKAGDCST-----AGTTE 1491  
QY 1553 VKLAKDLTLNLSVNAVGTGKIDKGISEVDANGQAKANTPVLSEANGLDLGGKVISNVGKT 1612  
DB 1492 TNVLDTIT-----GO-----LELG-----1505  
QY 1613 KDTDAANY-----OQLNEVRNLLGLGNDNAD-----GNQVNIADIKKOPNSGSSNRT 1660  
DB 1506 --ATTANVLFNDVAQLTQTGNIGGFLDFNAKNMGVTLNNNVNAGAVO--NTGGTNGT 1561  
QY 1661 VIKAGTVLGGKGNNDTEKLA---TGGVQGVVDKDGDNAN--GDLSNVWVKYTKDGSKKALLA 1716  
DB 1562 LI-----VLGASHNLNRYNGIAMLKVGAGNVITIAKGGKIGEIOGTGNT-----LTPA 1611  
QY 1717 TYNAAGOTNYVTNPAEALDRINEQGIREFHVNDGNOEPVQGRNGIDSSASGRKHSVAIG 1776  
DB 1612 HENLTG-----SINKTGGQALKLNFPMNGSVSCV-----1641  
QY 1777 FQAKGDEAAVAIGROTOAGNOSIAGDNAQAT-----GDQSI-----IGTGNVWACKHSGA 1829  
DB 1642 -----GTANSVGDTITTAGATSFASSVNAKCTATLGGTTSFANTFTTGAVTIAK-----1691  
QY 1830 IGDPTVKADNSYSVGNNOQFDTQTDVFGVGNNTITTESNSVALGSNSAISAGTHAGT 1889  
DB 1692 -----GSITSFAKNVTATSF--VANSATINFSLSAFNSN--ITGG-----1728  
QY 1890 QAKKSDGTAGTITTTAG-----TCTVKGFAGQTAAGVAVSVA--SGAERRIONVAAGEVS 1942

Db 1729 -----GTTTLGACQVYTG-TGSFTDLTLNLTTFDGAAGSGNLIKSGSTLDLS 1778  
QY 1943 ATSTDAVNGSOLYKATQSTANATN 1966  
Db 1779 GVSTLAL-----VVTATN 1791

RESULT 2  
YDBA\_ECOLI  
ID YDBA\_ECOLI STANDARD; PRT: 2003 AA.  
AC P33666; P76087; P76088; P76856; P76857; P76859;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL 205.9 KDA PROTEIN IN PAAY-ACPD INTERGENIC REGION.  
GN YDBA OR B1401/B1405.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;

[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Kasei H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
RA Aiba H., Kashimoto K., Kimura S., Kitagawa M., Kitagawa M.,  
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
RA Nakade S., Nakamura Y., Nishimura H., Nishio Y., Oshima T., Saito N.,  
RA Sampei K., Seki Y., Sivasubram S., Tagami H., Takeda J.,  
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 28.0-40.1 min region on the linkage map.";  
RL DNA Res. 3:363-377(1996).  
RN [3]  
RN SEQUENCE OF 464-2003 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=92190338; PubMed=1665988;  
RA Moszer I., Glaser P., Danchin A.;  
RT "Multiple IS insertion sequences near the replication terminus in  
Escherichia coli K-12.";  
RL Biochimie 73:1361-1374(1991).  
CC -!- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYSG (AC P25928).  
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR  
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT  
BETWEEN AMINO ACIDS 839 AND 840.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000237; AAC74483.1; ALT\_SEQ.  
DR EMBL; AE000237; AAC74487.1; ALT\_SEQ.  
DR EMBL; D90778; BAAL15009.1; ALT\_SEQ.  
DR EMBL; D90778; BAAL18880.1; ALT\_SEQ.  
DR EMBL; D90779; BAA18881.1; ALT\_SEQ.  
DR EMBL; X62680; -; NOT\_ANNOTATED\_CDS.  
DR EcoGene; EG11307; ydbA.  
KW Hypothetical protein.  
FT CONFLICT 489 489 I -> V (IN REF. 2).

FT CONFLICT 495 495 I -> V (IN REF. 2).  
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;

Query Match 4.8%; Score 496; DB 1; Length 2003;  
Best-Local Similarity 20.6%; Pred. No. 8.9e-11;  
Matches 399; Conservative 227; Mismatches 663; Indels 646; Gaps 99;

QY 284 ILDRNNTDASVYPLGKTLADQ-YKATRGQDSTDFISGNSNNNNSSIRKINVGAGS 342  
Db 147 ILDKTEKT-----LTIKRSVFTYVENADGT--ISLQDSNGRKATI----- 184  
QY 343 RDTDAVVAQLKVEELANRKITFKGDCDNNNSVERGLGNTLTIKGDAQTALTEANIG 402  
Db 185 -----NLWQI-----DEANNTVALEGSADGATKWQYNHNGELVITGD---NATVNNKG 231  
QY 403 VYTDGNGLVKLAKELTGLTSVATNKITVNTNNAELQSGLTFSPITGKTDKTVY 462  
Db 232 TVVDG-----KDSGTG-TEIN-----GNKGVIQDGLDLVSG----- 261  
QY 463 SIDGLAFTNDSNIAATKGTTRIT-KKIGFAGTGDVDESKPYLDNEKLVGNSTLNSGS 521  
Db 262 GGHGIDITGDSATVDNKGTMVTPDPESMGIO-----IDGKAIVNNE---GESTINGG 312  
QY 522 LPVNTTGNKQIQGVANGIKFATVANNVANTSATVGTARITEEKIGFAGTNDGVDEQAPY 581  
Db 313 -----TGQT-----INGDD--ATANNGKT-----TVDGKDSGTGE 341  
QY 582 LDKERLKGVRVEITDTSGINAGNH--KITGLTNGIANTDAVTIKQLDKAKPTLNAGD--- 636  
Db 342 INGN--GKVIQDGLDVSGGGHGDITG-----DSATV-----DNKGTMTVTPDES 386  
QY 637 -GISIN-----SNGDLVDSDSNIITPTYNISVKTTLN-SNGTSGNNKFSVSNADNN 688  
Db 387 IGIVQDGDQAVVNEGESAITNGS-----TGQINGDDATANNGKTTVDGKDS- 435  
QY 689 SLVTAKDLADLYLNKVNEDATDALSALPSFKVQNGDNNNAITVKGDTNKTFTLKLKGENGV 748  
Db 436 ---TGTEIAGNGKV-----IQGD-----LDVSG-----GGHGI 462  
QY 749 NITNRTAT---GTVTFGIDQSNGL-----TTPKLTGVSPTNGNRLVIEQ 789  
Db 463 DITGDSATVDNKGTMVTPDPESIGIQDQALVNEGESTITNGGTGTOIGN----- 516  
QY 790 VPSADGNSTKNIIGLSPTLPSIAPSG-----RNIALGNTIEEKDK- 831  
Db 517 -DATANNSGKTTVDGKDSGTGKIAGNIGIVNLGSLTVTGAGHVENIGDNGTVNNKGI 575  
QY 832 --SNAASIDDLVNA-CFNLKNNCKDKDFVSTYDTVDIDGNATATVYDEANQTSKVAY 888  
Db 576 VVSDTGSIGVLINGEGATVSTGDNV-----VSNEATGFSITTSNGKSVLAGSM 624  
QY 889 DVNVDEKTELTGDNCKKOLGVKTIKLTETSTNGNATFTSTDDHALVKASDIAGNLNTL 948  
Db 625 QVGDPSGTVDLNGNNNSVTLAAKDLKVVOKATG-----INVSGDANTVNIT 671  
QY 949 ABEIHTTKTANTALQFTTVKKYVDENDKADDTNATVKG-----DGTSGKVNTLKLK 1001  
Db 672 GNVLVKDKTADNAAEYFFDPSPGVINVSGDNNVTLDGKLTVVSDSEVTSRQSNLF----- 727  
QY 1002 KNCGLDIKTDKDTGTVTGINTQSGLKA-GDSTTLNNG-----LSIKNT-ASNEQIQVADG 1055  
Db 728 -----DGSA-----EKTSLGVVIGDNTVMNGGLELIGEKNALADGSQVSLRTG 773  
QY 1056 VKF--AMVNVNGVVGAGIDGTTTRITD-EIGFTGTNGSLDKSKPHLSKDGINAGKKTINI 1112  
Db 774 YSTSVIVVSGESSVYVNGDITISGEPPLGAGV-----IRVODKALLEI 818  
QY 1113 QSCEIAKNSHDAVTGGKIYDLKTELENKISSTAKTQNSLHEFSVADEQNNFTVSNPYS 1172  
Db 819 GSG-----ATLTMQDIDDSFEHHG----- 836  
QY 1173 SYDTSKTSVDV--ITFA---GENGITTGVKNGVVRVVGIDQTKGLTTPKLTVGNNGK---- 1223

Db	837	-----TRTVEIQNLGEFAFVTGEN--TTGINSGLTSLSLONGKDPAPSPVILLATNGSGATNA	890	DT	30-MAY-2000 (Rel. 39, Last sequence update)
Qy	1224	GIVINSQNGNTI-----TGLSNTLANVTNDKGSVR--TTEQGNIIKDEKTRAASIVDYL	1277	DE	30-MAY-2000 (Rel. 39, Last annotation update)
Db	891	CTITCKVTEQHSVFNKYSTGTSNSFI--FNNDVSSITCLVAQSN-----STIINTD	939	DE	OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
Qy	1278	SAGNLOGNGEAVDFSTYDVFNFANGNTTAKVTYD--DTSKTSKVVDVNVDDTTIEV	1335	DE	(SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
Db	940	SGIIDLYGRG--SVGMLAIADSTAENQKTLDSMWVDANDTTAMRDIAANSALDFGT---	995	DE	[CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
Qy	1336	KDKLGVKTTTLSTGTCGANKFALSNGATDALVKASDIIVAHNLNTLSGDIQTAKGASQAN	1395	GN	(120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
Db	996	-----GVGVGSDTSFGAGKATATL--NLGG-----VITIYNAGAG--MAAYGAS---	1036	OC	OMP B OR SPA OR RP704.
Qy	1396	NSAGYVDADGNKVIYDSTDKYIQAQKNDGTVDKTKVEKAKDKLVAQAQTPDGTTLAQMNVKS	1455	OS	Rickettsia prowazekii.
Db	1037	-----NTVINQGTIN-----LEKNGYDGS--LAANTLVGMAYVEHGT-----	1072	OC	Bacteria: Proteobacteria; alpha subdivision: Rickettsiales;
Qy	1456	VINKEQVNDANKKOGINEDNAFVKGLEKAAASDNKTKNAATVVGDL-----NA	1502	OC	Rickettsiaceae; Rickettsia.
Db	1073	-----AINDQGTVININVTG-----QAFYNDGTGTIVNYGTICTFGVCQSGNEYNNT	1120	OX	NCBI_taxid=782;
Qy	1503	VAQPLTFACDGTGTAKKGLTTLIKGGQDTDNKLTON--NIGVYAGTDGFTVKLAKDLT	1560	RN	[1]
Db	1121	DDFTSLIYTG--GDTITRSGETVTLNKSAAVTDKLAGNVVNSGTLSG--DOITVS-----	1172	RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
Qy	1561	NLNSVNAGG-----TKDEKISFVDANGQAKANTPVLNGLDLGKGVISNYGK-----	1610	RC	STRAIN=BREINL;
Db	1173	GILLETSGGIIINLVKLDK-----GAVIKNAGVMT--NNVDVSGGILLNAGEMTAQI	1222	RC	MEDLINE=91045572; PubMed=2122457;
Qy	1611	-----CTKQDAAVQQLNEVRNLLGLGNDNADGNQVNTADIKKDPNSGSSNRTRVIKAGT	1666	RA	Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
Db	1223	TWAGADSSLVNTGTINKIVQNAVFN-----NSGSVTGRMW-----	1260	RA	"Characterization of the gene encoding the protective paracrystalline-
Qy	1667	VLGGRKNNDTEKLTAGGVQGVVDKNGDNGDLSNVW--VKTOKDGSKKALLATYNAAGOTN	1725	RT	surface-layer protein of Rickettsia prowazekii; presence of a
Db	1261	SAGGVFNQTD-----GAIMRGAALGTGTAVANEGTWNILGSSSEGNNTGMLEVNNSAFNN	1316	RT	truncated identical homolog in Rickettsia typhi.;
Qy	1726	-----YVTNPAEAIIDRINQGIREFHVDNGOEPVYVQGRNGIDSSASGRHSVAIGFQAKAD	1782	RN	Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
Db	1317	REFILINDKNNAV--HINSGTLY-----NIGHM-----NLSNHHGAVNMW-----	1357	RP	[2]
Qy	1783	GEAAVAIGRQTOAGNOSTAIGNAQAQDQOSTAIGTGNVAVACKHSGALGDFSTVKADNSY	1842	RC	SEQUENCE FROM N.A.
Db	1358	-----GGNGRFINDTGIDVSAKSLV-----VSANNA--	1383	RC	STRAIN=MADRID E;
Qy	1843	SVGNNOFTDATQTDVFGVGNV-----ITVTSNSVALGNSAISAGTHAGTQAQKSD---	1895	RX	MEDLINE=99039499; PubMed=9823893;
Db	1384	--GDQNAFFWQDNGVINFDHDSASAVKTRHSNFTA--QNDGIMNISGTGAVAMEGDKNA	1439	RX	Medline=92114896; PubMed=1370573;
Qy	1896	-----GTAGTFTT-----AGAT-----GTVKGFAGQTAVGAVSVGASGA	1929	RA	Ching W.M., Carl M., Dasch G.A.;
Db	1440	QLVNNNTINLTAGTGTDTGCMGOLDANATADAVIENNGTINIFANDSFASFV--LGTVGH	1498	RA	"Mapping of monoclonal antibody binding sites on CNR fragments of
Qy	1930	ERRIQNVAAGEVSATSDAVNSOLYKATQSI-----ANATNELDHRHONKANAGAGISS	1985	RT	the S-layer protein antigens of Rickettsia typhi and Rickettsia
Db	1499	-----VYVNGFWIADGTGSGLLTKQGDGSLNVEGMNGNNSSEVHGD-----	1542	RT	prowazekii.;
Qy	1986	AMAMASMPQAIYIPGRSMVTTGGI-----ATHNGOGAVAVGLSKLSDNGQWYFKNGSA---	2037	RL	Mol. Immunol. 29:95-105(1992).
Db	1543	-----YTLPDVPPKPTNVTSGDSGAGSMNLLNGYVVGNTN-----VNGSAGKLK	1587	RN	[3]
Qy	2038	-DTQGHVGAAGVAGGF 2051		RP	IDENTIFICATION OF CLEAVAGE SITE.
Db	1588	VNASMNGVVEINTGP 1602		RP	MEDLINE=92104668; PubMed=1723180;
RESULT	3			RT	Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;
OMP_B_RICPR				RT	"Evidence for proteolytic cleavage of the 120-kilodalton outer
ID	OMP_B_RICPR	STANDARD;	PRT; 1643 AA.	RT	membrane protein of rickettsiae; Identification of an avirulent mutant
AC	Q53020; Q92CM0;			RL	deficient in processing.;
DT	30-MAY-2000 (Rel. 39, Created)			RL	Infect. Immun. 60:159-165(1992).

1-1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.

1-1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.

1-1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----

EMBL; M37647; AAA26390.1; ALT\_INIT.

EMBL; AF161079; AAD4234.1; --

EMBL; AJ235273; CAA15140.1; --

Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.

CHAIN 1 1328

FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.  
FT VARIANT 257 257 Y -> A (IN STRAIN BREINL).  
FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).  
FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).  
FT CONFLICT 178 179 AA -> VC (IN REF. 1).  
FT CONFLICT 191 201 TQOEAPLTGA -> INSRSSSYHLVS (IN REF. 1).  
FT CONFLICT 212 212 T -> I (IN REF. 1).  
FT CONFLICT 313 313 Q -> L (IN REF. 1).  
FT CONFLICT 1104 1104 D -> G (IN REF. 2).  
FT CONFLICT 1123 1123 T -> S (IN REF. 2).  
SQ SEQUENCE 1643 AA; 169854 MW; 735PDE392E6346CC CRC64;

Query Match 4.4%; Score 456.5; DB 1; Length 1643;  
Best Local Similarity 22.4%; Pred. No. 1.8e-09;  
Matches 396; Conservative 227; Mismatches 628; Indels 517; Gaps 100;

QY 517 LNSGSLTVNNTGKNOIQVGANGIKF--ATVANNVANTSATVGTARITEKIGF---AGT 571  
DB 12 ISAGLVTAATAT-----IVAGFGVAMGAAMOYNRTTNAAT-----TFDGGFDQAAGA 61  
QY 572 NQVDQAPYLDKERLKVGRVEITDSG----INAGNHKITGLTINGIANTDAVTI----- 622  
DB 62 NIPV---AP-----NSVTANANNPITFNTPNGLHNSLFLDTANDLAVTINEDTT 108  
QY 623 -----KOLDKAKPTLNAAGDISINSGDLVDSSGNITPTPTYNISVKTKLNSNGTS 674  
DB 109 LGFITIAQAQAFNFNTVAAGKILNI--TQGGITVQEASN---TINAQNALTKVH----- 158  
QY 675 GNNKSVSNAHDNNSLVATKADLADYLNKYNETADSLPSFKVONGDSSNNAITVGKDTNG 734  
DB 159 -----GGAINANDLSG--LGSITFAAPSVLEFNLIINPTQAPLFLG---ANS 203  
QY 735 KTFNTLKLKXENG-----VNITNTRATGTVTFGIDQSNGL---TTP-----KLTVG 777  
DB 204 KIVN-----GGNGTLNITNGFIQVSDNTFAGIKTINIDCQGLMFNSTPDAANTLNLOVS 258  
QY 778 SDT---NG-----NRLVIEQVPSADGNSTK--NIKGLSPILPSIASPSGRNIALGNITIEK 829  
DB 259 GNTINFNGIDGTGKLVL---VSKNGAATFENVGTGLGGNLKGIELNTAAVA--GKLISQ 314  
QY 830 DKSAASIDVLNAGPNLKNKGDKDFVSTYDVFIDGNATTTAT-----VTYDEANQ 882  
DB 315 GAANAVIGTD-----NGAGRAAGFIVSD-----NGNAATISQVYAKNVMVIOASANA 361  
QY 883 TSKVAYDVNVDEKTIELTDGNGKKQLGVKTIKLTETSTNGNATFTSTDDHALVKASDI-- 941  
DB 362 GQOVTEHIVD---VGLGTTNFKTADSKVI--ITENS--NFGSTNFCNLQDTQIVVPDKIL 416  
QY 942 AGNLNTLAEIHTTKTANTA--LOFTVK-----KVDENDKADDTNATVVGKDGTP---S 991  
DB 417 KGNF-----IGDVKNNGNTAGVITENANGALVSASTDPNIAVTNINAIEAEGAGYVELS 470  
QY 992 GKVNTLKLKGNCLDIKTKDQGVTVFGINTQSLKAGDSTTLNNGLSIKNTASNEQIQV 1051  
DB 471 G-IHIAELRLNGGSLFKLADGTIVNG-----PVQN----- 501  
QY 1052 GADGVKFAVMNVNGVWAG---IDGTTTRITRDEIGFTGNSL-----DKSPHLSKDG 1101  
DB 502 -----ALMNNALAGSITLDGSAIITGD--IGNGVNAALQHIITLANDASK--ILALDG 552  
QY 1102 INAGGKTYNIQSGETAKNSHDAVTGCKI-----YDLKTELENKISSTAKT 1147  
DB 553 ANIGANV-----GGAT-----HFOANGGTIKLTNTQNNIVVNFOLDITTDKTVGVVDASSET 604  
QY 1148 AQNSLH--EFSVADEQGNFTVSNPYSSYDTSKTSIDVTITFAGENGITTKVNGVVRVGIQ 1206  
DB 605 NQOTLINGSIGTVVANTKTLAG--UNIGSSKT---ILNAGDVAINELVNIENNGSVQLNH 659  
QY 1207 TKGLATPKLTGVNNGNGKGVINSQNGQNTITGLSNTLANVTNDKGSVRTTEQGNIIKDED 1266  
DB 660 NTYLIKTINAAN---QGQIIAADPLNTNT---FLADGTN--LGSA-----ENPLSTIHF 707

## RESULT 4

OMP\_RICJA STANDARD; PRT; 1656 AA.  
ID OMP\_RICJA AC 006653;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE OUTER MEMBRANE PROTEIN B PRECURSOR (166 KDA SURFACE-LAYER PROTEIN)

QY 1267 KTRAAISIVDVLSA--GFNLOGNGEAVDFVSTYD-----TVNFANGNTTTAKVTY--DDTSK 1318  
DB 708 ATKRAADNSILNVGKGVLYANN-----ITNDANVGSLSHFRSGGTIVSGTGGQOCHK 762  
QY 1319 TSKVVYDVNVDDTTIEVKDKKLGKVTTLTSTGTGANKFALSNOATDALVKASDIVAHL 1378  
DB 763 LNNLILD---NGTIV---KFLGDTTFNGGKTIEGKLSILQISNNYTPDH--VESADNTGTL 814  
QY 1379 NTLSGDIOTAKGASQANNSAGY-----VDADGNKRIYDSTDNDKYYQAKNDGTVDKTK 1430  
DB 815 EFVNTDPITV---TLNKQAGFGLVKQVIISGPN--IVFNEI-----GNVGIVH 859  
QY 1431 EVAKDKLVAAQAQT-----PDGT--LAQMNVKSVINKEQVNDANKKOGINEDNAFVAGLE 1482  
DB 860 GIAANSISFENASLGTSLSFPGTPLDVLTIKTKVTGNGTVDN-----FNAPIVVWSGID 913  
QY 1483 KAASD-----NKTNAAVTVGDLNNAVAQTPLTFAGDTGTTAKKLGETLTTKGGQ----- 1531  
DB 914 SMINNGIIGDKKKNIILALSIGSDNSIIVNANTLYSGIRTTKNNQG--TVILSGGPNPNPGT 972  
QY 1532 -----TDTNKLTD---NNI-----GVVAGTDGFTVKLAKDL 1559  
DB 973 IYGLGLENGSPKLGQVTFTTDDYNNLGSIIANNVTINDYVTLTTGGIAGTD--FDAKI--- 1027  
QY 1560 TNLNSVNAGCTKIDEGKISFVDAN-----GQAKANTPVLSSANGLDLGKVVISNVG 1609  
DB 1028 -TLGSYNGNA-----NVREVDSTFSDPRSMIVATQANKGTVTY-----LGNALVSNI- 1073  
QY 1610 KGFKDDAANVQOLNEVNRLLGLGNDNADGNOYNIADIKKDPNSGSSSNRTVTKAGTVLG 1669  
DB 1074 -GSLDTPVASVR-----FTGNDSGAGLQGNITYSQNIID---FGYNTILNSNVILG 1120  
QY 1670 GK-----DTEKLATGGV--QGVVDKDG----- 1692  
DB 1121 GGFTAINGEIDLTLNLIIFANGTSTWGDNTSISTTLNVSSNGIGVVIAEDAQVNAATTG 1180  
QY 1693 -----NANGDLSNVVKTQDKGSKALLATYNA--GQFNVTYNNPAAIDRINQGI 1743  
DB 1181 TTTIKIQDNANANFSGTQATYTLIQG-----ARFNGTLAGPNAFVGTG-----SNIFV 1227  
QY 1744 REFHVNDGNOPEPVQGRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNOSTAIG 1803  
DB 1228 KYELIHDSNODYVLTTRTNDVNLVVT-----TAVGNSAIFA-- 1261  
QY 1804 DNAQATGDQ--SIAIGTGNVYVAGKHSIGAIGDPS-----TVKADNSYSVG-----NNNQFT 1851  
DB 1262 -NAPGVSONISRCLESTNTAAAYNNMLLAKDPSDVATPVGAIAITDTSAAVTVNLNDTQKT 1320  
QY 1852 DATQTDVFGVGNITVYESNSVALGNSAISAGTHA---GTOAK-----KSDGTA 1898  
DB 1321 QDILLSNRLGTLRYLSNAETSDVAGSATGAYSSGDEAEVSYGVNAKPFYNTAEQDKGGIA 1380  
QY 1899 GTTTTAGATCTVAG---FAGOTAVGAVSVGASGAERIQHVAAGEVSATSTDVANG--SQ 1953  
DB 1381 G--YKATGTGVVGLDTLASDNLMLGAAIGITKTDIKHQDYKKGD-----KTD--INGLSFS 1433  
QY 1954 LYKATOSIAN-----ATNELDHRHONENKANAGISSAMAMASMPQAYIPGRSNVT 2004  
DB 1434 LYGSQQLVKNFAGNAIFTLNKKVSKSORFYFESNGKMSKOIAAGNYDNMTF--GGNLIF 1492  
QY 2005 GGIATHNGQCAVAV---GLSKLSDNCO 2028  
DB 1493 G--YDYNAMPNVLVTPMAGLSYLKSSNE 1518

DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (ROMP B)  
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
GN OMPB.  
OS Rickettsia japonica.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=35790;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=YH;  
RA Uchiyama T.;  
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia  
RT japonica";  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
CC SIMILARITY).  
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
CC (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
CC LAYER WITH HEXAGONAL SYMMETRY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB003681; BAA20138.1; --  
DR Antigen; S-layer; 120 KDA SURFACE-EXPOSED PROTEIN.  
KW CHAIN 1 1338  
FT CHAIN 1339 1656  
FT DOMAIN 528 533 POLY-GLY.  
FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;  
SQ  
  
Query Match 4.4%; Score 455.5; DB 1; Length 1656;  
Best Local Similarity 22.0%; Pred. No. 2e-09; Indels 533; Gaps 96;  
Matches 397; Conservative 204; Mismatches 673;  
  
Qy 490 GFAGTNGVDKPYLDNEKLVKGNSTLNSGSLTVNNTTC-NKQIOVGA-NGIKFATVAN 547  
Db 45 GVATVDOVG-----PDQVALANVAVAPNAVITANANNINGINLNTAGSFNGL-FLSNAN 98  
Qy 548 NVANTSATVGTARTEKEKIGFAGTNGVDQAPYLDKERLVGRVEITTDGSGINAGNHKI 607  
Db 99 NLAVT-----VSE-----DTTLGFINNAANANRNLTLDAKTL---TI 135  
Qy 608 TGLNGTANTDAVTIKOLDAKPTLNAGDGLISNSNNG-DLVDSSGNITPTPTNIVSKTT 666  
Db 136 TG--QGITNVSQAATHRAQNIKAVFNGGAALANDLSGLGTIDFGAAASTLVFDLANPTT 193  
Qy 667 K-----LNSNGTSGNKKFVSNAHDNNSLVTAADLADLYNKVN---ETADSAALPS 713  
Db 194 OKAPLILADNALIVNGANGT-----LNVNNGFIQVSDKSFAT 230  
Qy 714 FRVQNGDNNNAITVGK-----DTNGKFTNLTLLKNGENYITNTRATGTVFGIDQSN 767  
Db 231 VK-----AINIGDQGFENFTNATNALNLQA-GGTTINFNCTDGTGRVLVLSKN 280  
Qy 768 GLUTPKLVGSTNGLRVICQVPSADGNSTKNIKGLSPLTLPISPSGNIALGNTIE 827  
Db 281 GAAT-----DFN-----VTGSLGNLKGIE-----LNTVAI-NGOLIA----- 313  
Qy 828 EKDKSNAASIDDLVLAGENLKNNGDK--DFVSTYDT--VDFIDGNATFATVYDEANQT 883  
Db 314 ----NAGPANAVIGT-----NNGAGRAAGFVVSVDNGKAATIDQGVYAKDMVIQSANAN 363  
Qy 884 SKVAY-----DVNVDEKTIELTGDNGKKQLGVKTIKLTETSGNATTTSTDDH--ALVK 937

Db 364 GQVNERHIVDVGIDGTTAFKT-----AASIVAITQNSNFG-----TTDFGNLAQVTT 410  
Qy 938 ASD-----IAGNLNTLAEELHTTKGTANTALQ--TPVVKVVDENKADDDTNAITVGRKDGTSK 993  
Db 411 VPDNTMLTGNFTGDANNPNTAGVITPAANGTCLASADANAV--TNNTAIE--ASG-- 465  
Qy 994 VNTLKLKNGKGLDITKDKDGTFTGINTQSGKAGDSTTLNNGLSIKNTASNEQIOVGA 1053  
Db 466 VGVVOLSGTHTAELRLGNAGSV-----FKLADGTVINCK-----VNOTVILVG-- 507  
Qy 1054 DGVKPFAMVNNGVVAG-----IDGTPRITRDEIGFTGTNGSLDKSPHLSKDG---INAGKK 1107  
Db 508 -----GVLAAGAITLDGSATITGD--IGNGGGGAALQSIT--LANDATKTLTLGGA 554  
Qy 1108 KTNIOSGEIAKNSHDAVTGGKIYDLKTELENKISSTAKTAQNSL-----HEPSVADEGNN 1164  
Db 555 NIIISANGGTI-----NFQANGGTI-----KLTST-----QNNIVVDCDLAIATDQTV 597  
Qy 1165 FTVSNPYSDYTSKTSDDVTFAGENGITTKVNGVVRVGDQTK-----GLTPKLTGVNN 1220  
Db 598 VDASS-LTNAQTLTISGTIGIIGANNITL-----GQFNIGSSKTTLNGGNVAINELVIGN 652  
Qy 1221 N-----GKG-IVINSQNGQNTITGLSNTLANVTNDKSGVTRTEQONI 1261  
Db 653 GSVQFAHNTYILTRTNAGQKGIIFNPVNNNTTLAAGTNLGSAAANPLAEINFGSKG-- 710  
Qy 1262 IKDEDKTRAASIVDVLSAGFNLOGNGEAVDFVSTYTVNFAN-----GNTTAKV 1311  
Db 711 -----ARADTVLN-----GEGVNLATNITTTDANVGSFVFNAGGKNIVSGTV 754  
Qy 1312 TYDTSKTSKVYDVNVDDTTIEVKDKLGVKTTTLTSTGTGANKFALSNQATGDALVKA 1371  
Db 755 GQOQNGKNFTVALD---NGTIV---KFLGNATFNNGNTTIAANSTLIQISGNVTADFIASA 807  
Qy 1372 --SDIVAHNTLSGDIQTAKGASOANNSAGY--VDADGNKVIYDSTDNKYQAKN-DGTYD 1427  
Db 808 DGTGIVEFVNTGPIINVTLNKQAVFVNALKOITVSGPGNVVV-----NEIGNAGNYIGAMT 862  
Qy 1428 KTEKAVADKLVQAQATPD-----GTLAQMNVKSVINKQVNDANKKQGINEDNAFVKG 1480  
Db 863 DTIAFENSSLGAVLFLPSGIPFNDAGNTIPLTIKSTVGNETA-----EGFSPSVIVSG 916  
Qy 1481 LEKAASDNKT---KNAAVTVG---DLNVAQTPPLTFAGDTGTAKKLGELTTLTKGGQD 1534  
Db 917 VDSVIADQGVIGDQNNITVGLGSDNGIIVNATTLYAG--IGTINNQG--TIVLSGGVPNT 974  
Qy 1535 NKLTDDNIGVVAGTDGF--TVKLAKDLTNLSVNAAGTKIDE-----KGISFVDA 1582  
Db 975 FG-TVYGLGTIGASKFKQVTFITDYNLNLGNIATNTINDGVTVTVTGGIAAGGIAGTDF 1033  
Qy 1583 NGQ-----AKANTPVLSANGL-----DLGKVISNVGKCTKDDTAA 1618  
Db 1034 DGKITLGSVNGNANVRFADGIFSNSTSMIVTTKANNGTFTVYLGNAFVGNI--GOSDTPVA 1091  
Qy 1619 NVQQLNEVRNLLGLGNDNADGNVNIADIKKDPNCGSSNRTVIKAGTVLGK----- 1671  
Db 1092 SVR-----FTGSNNGAGLKNLIYQVID---FCYVNLGIVNSNVILGSGTAINCK 1139  
Qy 1672 -----GNN-----DTEKLTG--GVQGVGDKDG-----N 1693  
Db 1140 IDLLTNTLTFAGTSTWGNNTSIETTLTLANGNIGHIVIAEAGQVNAATTTGTTINVDN 1199  
Qy 1694 ANGDLSNVWVKTKQKSKK-----ALLATYNAAG 1722  
Db 1200 ANANFSGTQTYTLIOGGARFNGLGPNFTVTSNRFVNYGLIRAANDYVITRTNNAEN 1259  
Qy 1723 -QTNVYTNPAEAIADRINEQIRFFH--VNDGNQEPVQVQNGIDSSASGKISVAIGFOAK 1780  
Db 1260 IVTNDITNSPFGGAPGVQGVNTTFVNATNTAAYNNLLAKNSADSANFVGTIVTDTSAI 1319  
Qy 1781 ADGEAAVAGTQAGNQSTAGDQNAQTGDSIAITGCTGNVAGKHSIGAIGDPSTVKA-- 1838





Thu Sep 13 14:19:31 2001

Db 804 TKRGS-QTEVAGDLTITANK-----DLLEHGAHHVEGRYQESGENIQHLAVNDS-- 853  
 Qy 1301 FANGNTTTAKVYDDTSKTK-----VYDVNDVTIEVKDKL---GVKTTTLTSTGTG 1353  
 Db 854 -----ETSKDLSNIGDVGNDYSGVTKPKRAIEDGVNTT-----KP 893  
 Qy 1354 ANKFAISNOATGDALVKAASDIIVAHNLTL-----GDIOTAKGASQA--N 1395  
 Db 894 GNNTDLTKKVT-----ARDIAIANLANLSNLETPNVGVEGIKGGGQQSQSDTSQAVSTS 947  
 Qy 1396 NSAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKVAKDKLVAAQATPGTL----- 1448  
 Db 948 INAGKIDIDSNNKLDH--OGTHYQSTQEGI-----SLTANTHTSEATLDKHKQTF 995  
 Qy 1449 -----AQMNVKSVINKQVNDANKKOGINEDNAFVKGLEKAASDNKTKAAVTVGD-- 1499  
 Db 996 HETKGGQIGVSTKGSQ--ITVAIKGEGQTTDNALMETKAKGSOFTSNGDISINVGNAH 1054  
 Qy 1500 -----LNVAQAQPLTFAGPT-----GTTAKKLGETLTIK-----G 1529  
 Db 1055 YEGAQFDAQKGTIVNAGGDLTLAQATDTHSESOSNVNGSANLKVGTTPESKDYGGGFNA 1114  
 Qy 1530 QOTDINK-LTDNNGVAGTGTGFTVKLAKDLT-----NLASNAGGTKID-- 1573  
 Db 1115 GTHHSKEQTTAKVGTITGSGQIEINAGHNLTLQGTLLSSEQDIALNATN---KVDLQS 1170  
 Qy 1574 -----EKGISF---VDANGQAKANTPVLSANGLDLGKVISNVGKGTQKTDAAVQOQLN 1624  
 Db 1171 ASSEHTEKGNLGGVAGFGKMTDDASSVNL---GSAQFAIGKQDEKSVSRGGGTIN 1227  
 Qy 1625 EVRNLLGLND-NADGNOVNIADIKKOPNSG---SSSNRTVIKA--GTVLG--GKGNND 1675  
 Db 1228 NSGNLTINGNSVHLOGAQVNSKDTLTQSOGDIEITSQAQTDYKKNWGDIGFNGKKTNN 1287  
 Qy 1676 T-----EKLAT-----GQVQGVGD-----KGNANGDLSNVWVKTKD-----GSKKA 1713  
 Db 1288 TPKEVTEEPATSIHNLGGKLLVNDVDOOKTSHQATLETGLTINSKNKDLTSLGANVTA 1347  
 Qy 1714 LLATYNAAGQTVNTPNPAEADIRINEQIR--FFHVDNGOEPVVGVRNGRIDSSASG-- 1769  
 Db 1348 DSVTGNVGSLSNTASQES---DRHVTGVNVGVNITNDPKSSQVNYK-----TAKAGGSL 1399  
 Qy 1770 -----KHSVAIGFQAKADGAEAAVAGROTAQGNOSTA-----TGDNAQATGDSIA-IGTG 1819  
 Db 1400 LEKTIKDTIDSGIKSTED-----AISKYNSLSSTIADTKTIGSDTKAKIDQGFQVGVNG 1454  
 Qy 1820 ---NVVACKHSGAIGDPSTVKADNSYSVGNNGNOFTDATQTDVFCVGNNTVITSNVALGS 1877  
 Db 1455 IKNIVTG-----AEGHTANADIKVTHVDNDVTKTT-----SLTSNNDLSLNV 1497  
 Qy 1878 NSAISAGTHAGTQAKSDGTAGTITTTAGATGTVKGFAGTAVGAVSVGASGABERRIQNVA 1937  
 Db 1498 N-----GSEKLTGA--EIVSQOGVDLGGSSV-----KLENIE 1528  
 Qy 1938 AGEVSATSDAVNGSOLYKATQSIAN-----ATNE 1967  
 Db 1529 GIHYEAGADLDLKSVDLAKOLQVGGDLSEKSPVKTKNE 1566

RESULT 6  
 OMPB\_RICKRI ID OMPB\_RICRI STANDARD; PRF: 1654 AA.  
 AC Q53047;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
 DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
 GN OMPB.  
 OS Rickettsia rickettsii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsia; Rickettsia.  
 OX NCBI\_TaxID=783;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R;  
 RX MEDLINE=92167802; PubMed=1724278;  
 RA Gilmore R.D. Jr., Cleplak W. Jr., Policastro P.F., Hackstadt T.;  
 RT "The 120 kilodalton outer membrane protein (rOMP B) of Rickettsia  
 rickettsii is encoded by an unusually long open reading frame:  
 evidence for protein processing from a large precursor.";  
 RL Mol. Microbiol. 5:2361-2370(1991).  
 RL [2]  
 RP SEQUENCE OF 279-1654 FROM N.A.  
 RC STRAIN=R;  
 RX MEDLINE=90136087; PubMed=2515418;  
 RA Gilmore R.D. Jr., Joste N., McDonald G.A.;  
 RT "Cloning, expression and sequence analysis of the gene encoding the  
 120 kD surface-exposed protein of Rickettsia rickettsii.";  
 RL Mol. Microbiol. 3:1579-1586(1989).  
 CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 LAYER WITH HEXAGONAL SYMMETRY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@sib-sib.ch).  
 CC  
 CC ENBL; X16353; CAA34403.1; -;  
 DR Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.  
 KW CHAIN 1 1333 32 KDA BETA PEPTIDE.  
 FT CHAIN 1334 1654 32 KDA BETA PEPTIDE.  
 FT DOMAIN 1181 1188 POLY-THR.  
 FT DOMAIN 1654 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;  
 SQ SEQUENCE

Query Match 4.38; Score 443.5; DB 1; Length 1654;  
 Best Local Similarity 22.04; Pred. No. 5.2e-09;  
 Matches 383; Conservative 189; Mismatches 567; Indels 605; Gaps 92;  
 Qy 650 SSGNITTPYNI-----SVKTTKLSNGTSGNNKFSVNAHDNNSLVTAADLYLNKY 703  
 Db 13 SAGLVASTATIVASFAGSANGAAIQNRRT-----NGAATTVDGAGF----- 55  
 Qy 704 NETADSLPSPKQVNGDNNNAITVVGKDTNCKTFNT-----LKLKGENGVNITTNRAT 756  
 Db 56 DQTAAPA-----NVGVNALNAVITANANNNGINENTPAGSFGNGLLNTANNLAVTVSEDT 108  
 Qy 757 --GTTFEGIDQSNGLTTPKLTVGSDTNGNRLV1--EQVPSADGNSTKNITKGLSPTLPSI 812  
 Db 109 TLGFTITNVVHNAHSF---NLTLNA---CKTLTITGOGVTNAQAATKN----- 150  
 Qy 813 ASPSGRNTALGNTIEEKDKSNAASIDDLVNLNAGFNKNGKDKDFVSTYDVFIDGNATT 872  
 Db 151 ----AQNVVV-----QFNNGAAIDN-----NDLKGVGR-----IDFGAPA 181  
 Qy 873 AIVTYDEANQTSKVAVDVNVDEKTIETGDNCKKOLGVK-TIKLTETSTNGNATTFSTDD 931  
 Db 182 STLVLNLANPT-----OKAPLIIGDNAVIANGVNGLNVTNGFIQVSNKSFAT-- 230  
 Qy 932 DHALVKASDIAGNLNTLAEIHTTKGTANTALQTTFTVKKVDENDKADDTNATITCKDCTS 991  
 Db 231 ----VKAINLADQGLII---FNTDANNANTL-----NLOAGGTTINFTDGTG 272  
 Qy 992 GKV-----NTLKLKGNGLDIK-----TDKDGTVT-----FGINTOSGLKA 1027  
 Db 273 RLVLKLSKHAATAATNFNTITSLGGLNKGKLVIEFNITVAVDQGLTANAGAANAVIGTNNGAGRAA 332



Db	624	PFIAGSOALLKQALNNKNNPFYAYYKQKGLTALTDPLKTVEMNTAQPIINDIYNVVI--VS	682
Qy	683	NAHDNNSLVITAKDLADYLNKVNNET--ADSAFPSKVFQNGDGNNSNNAITVQKDTNGKTFNPL	740
Db	683	PRQAGLVVDKAAIDALEKNPSTVVAENGYPAVELKDLFTSD-----KTF-----	728
Qy	741	KLKGENGVNITTRNATGTVTFGIDQSNGLTTPKLTIVGSUTNGNRLVIEOVPSADGNSTKN	800
Db	729	KL-----TFTNRTHLTYQMD-SNTDTPNAVYTSATDPNSGVLYDKKIDGA-----	773
Qy	801	LIKLSPTLPSIASPSGRNIALGNITIE-EKDKSNAASIDDLVN-----AGFNLK-----	848
Db	774	AIKAGS-----NITVPACKTAQIEFTLSLPKSFQDQOQFVEGLFNKFGKDSRLNLPYMGFF	829
Qy	849	-----NNGKDKFVS-----TYDVTDFIDGNATFAT-----VNYDEANOT---SKVA	887
Db	830	GDWMDGKIVDSLINGITYSPAGNGFTVPLLT-NKNTGTQYGGMVTDAOGNQIVDDQATA	888
Qy	888	Y-----DV-----NVDEKTTLTGDNKKQKLGKVTIKLTETSTNGNATTFSTD	930
Db	889	FSSDKNALYNDISMKYYLLLRNISNVQVDILDGQGNK-----VTLSSSTNRKKTYYNAH	942
Qy	931	DPHAL-----VKASDIAGNLTUAEELHTTKGTANTALQTFYKVKDE--NDKAD-----	978
Db	943	SQOYIYHAPAWD-----GTYYDQRDGNIKTADGGSYTYRISGVPEGDKRQVDFVPFKL	997
Qy	979	DTNATTVGKDGTSQKVNLTCLKGKNGLDIKTDKDC-----TVTFEGINTQSLKA--GDST	1031
Db	998	DSKAPTIVRHVALSAKTENGKTYLYLTAEKADDLGLDGLATKSVKTAINEVTLNLDATFTDAG	1057
Qy	1032	TLNNGLSLTKNTASNEQTOVGADGVKFAWNVGVGAGIDGCTTRITRDEIGFTCTGSLD	1091
Db	1058	TADGTYKLETLSLSDQQA-----LGNG-----DMSAELYLDNASNATDQDAS	1102
Qy	1092	KSKPHLSKDG--INAGG--KKITINQSGEIAKNSHDAVT-GCKYIDLKTELENIKLSSTAK	1146
Db	1103	VQPGSTFSDLVINGGGIPDKISSITTG-----YEANTQGGTFTFSGTYPAAVDGTYT	1156
Qy	1147	TAQNSLHESVA--DEOGNFTVSNPYSSYDTSKTSVDITFACENGI---TTKVNKGVVVR	1201
Db	1157	DAQCKKHDLNTTYDYAATNSFTASMPVTNADYAQVDLTADKHAHTQLKHFDTKVR-----	1211
Qy	1202	VGIDQTKGLTTPKLT--VGNNNKG-----GIV-----INSQNGQNTITGL-----	1239
Db	1212	-----LTAPTFTDLKFNNGSDQTSSEATIKVTGTVSADTFTVNVGDTVAALDAQHHFS	1263
Qy	1240	-----SNTLANVINDKGSVPTTEQGNLIKDEKTRAASIVDVLISAGFNLQGNCEAVD	1291
Db	1264	VDVPVNYGDRTIKVTAFDEGNTTTEQKLTSSYDP-----DMLK-----NSVTFD	1309
Qy	1292	FVSTYDTPVNEFANGNTTTAKVTYDDTSKTSKVYVDVNVDDTTEVKKDKLGVK-----	1346
Db	1310	QGVTFGANEF---NATSAKE-YDPKGTGIAITLTKGVKHPTTTTLQVQDGKQIPKDDLTFSFT	1365
Qy	1347	LTSTGTCANKFAL-----SNOATGDALVKASDIVAHLNLTLSGDIOTAKGASQANNAGY	1400
Db	1366	LDLGLTLGCKPFGVVVGDDTQNKTFQEALTFILDAVA--PTLSLDSST-----DAPVY	1415
Qy	1401	VDADGNKVIYDSTDN-KYYQAKNDGT-----VDKTEKVAKDKLIVAAQATPDGTLQAMNV	1453
Db	1416	TNNPNFQITGATDNAQYLSLSINGSSVAQYVDININSGKPGHMAIDQ-PVKLLLEGKNV	1474
Qy	1454	KSVINKEOVDANKKQGINEDONAFVKGLEKAASDNKTKNAAVTVGDLNAVAQPTPLTFAGD	1513
Db	1475	LTV-----AVTD-----SEDNTTKNITVYEPKTKL-AAPTIV-----TP-----	1508
Qy	1514	TGTTAKKLGETLTIKGGOTDNKLTNNNIGVAGTGDGFTVKLAKDLTLNLSNVNAGG--T	1570
Db	1509	STTEPAKTVTLTANSAAT-----GETVOYSAD-----GGKTYQ	1541
Qy	1571	KIDEGISFVDANGAOKANTPVLISANGLDLGKVIISNVGKTKDTKDIDANVQOL-NEVRNL	1629

QY 1630 LGLGNDNADGNQVNIADIKDPNSGSSNRRTVIKAGTVLGGKGNDEKLTATG--GVQVG 1687  
 Db 1597 -----IASAKTSLASGKYDDATT-----TALAAA-----TQRAQTALDQTNAS 1634  
 QY 1688 VDKDGNANGDLSNVVYKTKDCKSKALLATYNAAGTQNVVTPNPAEIDRINEOGIRFFH 1747  
 Db 1635 VDSLTCANRDLQT-----AINQLAKLPADKKTSL--NQLQSVKAALETDL---- 1679  
 QY 1748 VNDGNOEPVVOGRNGIDSSASGKSHVAIGFOAKADEAAVAIGROTQAGNOSTAIGDNAQ 1807  
 Db 1680 ---GNO-----TDSSTGK-----TFTAALDDLVA-----QAQACTQT---DDQLQ 1713  
 QY 1808 ATGDQSIATGTCNVVAGKHSIGAIGDPSTVKADNSVSGVNNQFTDATQTDVFGVGNNTIV 1867  
 Db 1714 AT-----LAKVLDVLAFLAKLAEG-----TKAATPAEVGNK-----DAATGKTWYADIADTL 1759  
 QY 1868 TESNSVALGSN-----SAISACTHAGTQAKSDGTAGTGTGTTTAGATGTVKGFAG 1915  
 Db 1760 TSGQASADASDKLAHLQALQSLTKVAVAEAKTVVKGDDGTGTGSDKGGGQGTAPATPG 1819  
 QY 1916 QPAVGAVSYGASGAEIRIQNVAAVEY-----SATSTD 1947  
 Db 1820 -----DIGKDKGDEGSPSSGGNIPTNPATTSTSTD 1851  
 RESULT 8  
 ID OMPB\_RICTY STANDARD; PRT; 1645 AA.  
 AC P96989;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
 DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
 GN SLP.  
 OS Rickettsia typhi.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 ON NCBI\_TaxID=785;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WILMINGTON;  
 RX MEDLINE=94040787; PubMed=8224886;  
 RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;  
 RT "Cloning and sequence analysis of the gene encoding the crystalline  
 RT surface layer protein of Rickettsia typhi";  
 RL Gene 133:129-133(1993).  
 RN [2]  
 RP PARTIAL SEQUENCE.  
 RC STRAIN=WILMINGTON;  
 RX MEDLINE=92114896; PubMed=1370573;  
 RA Ching W.M., Carl M., Dasch G.A.;  
 RT "Mapping of monoclonal antibody binding sites on CNR fragments of  
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia  
 RT prowazekii";  
 RL Mol. Immunol. 29:95-105(1992).  
 RN [3]  
 RP IDENTIFICATION OF CLEAVAGE SITE.  
 RX MEDLINE=92104668; PubMed=1729180;  
 RA Hackstadt T., Messer R., Cleplak W., Peacock M.G.;  
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer  
 RT membrane protein of rickettsiae: identification of an avirulent  
 RT mutant deficient in processing";  
 RL Infect. Immun. 60:159-165(1992).  
 CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.  
 CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A  
 CC S-LAYER WITH HEXAGONAL SYMMETRY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; L04661; AAB48987.1; -  
 KW Antigen; S-layer; Transmembrane; Cell wall.  
 FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT CHAIN 1354 1645 32 KDA BETA PEPTIDE.  
 FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).  
 FT CONFLICT 657 657 H -> N (IN REF. 2).  
 FT CONFLICT 842 842 V -> I (IN REF. 2).  
 FT CONFLICT 1071 1071 G -> A (IN REF. 2).  
 FT CONFLICT 1306 1306 G -> S (IN REF. 2).  
 SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;  
 Query Match 4.2%; Score 436.5; DB 1; Length 1645;  
 Best Local Similarity 21.8%; Pred. No. 9.2e-09;  
 Matches 366; Conservative 185; Mismatches 566; Indels 563; Gaps 83;  
 QY 650 SSGNITPTPTNISV-----KTKLNSGTSGNKFYSNAHDN-----NSL 690  
 Db 13 SAGLVASTATIVAGFSGVAMGVMQVNRNTTNAATVTDGAG-FDQTGAGVNLPAVNSV 71  
 QY 691 VTAKLADLYLNKVNETHADSALPSFKVQNGDNSNNAITVGDKTNGKTFNTLKLGENGVNI 750  
 Db 72 ITA-----NSNNAITF--NTPNGNLNLSLDTANTLAV 102  
 QY 751 TYNRAT--GTVPFGIDQSNGLTTPKLTVGSDTN---GNRLVIEQVPSADGNTKNIKG 804  
 Db 103 TINENTLGFVNTVKQGNFF---NETIGAKSLTITGHGITAQQ--AATKSAQNVS- 156  
 QY 805 LSPTLSIASPSGRNIALGNTIEEKDKSNAASID---DVLNAGFNLN----- 849  
 Db 157 -----KVNAGAAIINDNDLSGVGSIIDFTAAPSVLE--FNLIPTTQEAPLTL 200  
 QY 850 -----NG-----KDKDFVSTYDTVFIDGNATTAIVTYDEANQTSKVAY 888  
 Db 201 GDNAKIVGANGILNITNGFVKVSDKTFAG-ITKINIGDQGLMFTNTPDAAN-----AL 254  
 QY 889 DVNVDEKTIELTDGNGKQOLGVKTIKLTETSTNGNATTFSTDDHALVKASDIAGNLNTL 948  
 Db 255 NLOGGNTINFNRDG-----TGKLVLSKNGNATEFNV-----TGSIGGNLKV 299  
 QY 949 AE-----EIHTTKGTANTALQT-----FTVKYVDENDKADDTNAITVGDGTS 991  
 Db 300 IEFDTTAAAGKLIANGGAANAVIGTDNGAGRAAGFIV-----SVDNGNAATI-----S 347  
 QY 992 GKVTNLKLGKNGLDIKTDKDGTVTF-----GINTQSLKAGDSTTLNNGLSIKNTAS 1045  
 Db 348 GQV-----YAKDIVIOSANAGGQVTFEHLVDVLGGKTNFKTADSKVI-----ITENAS 396  
 QY 1046 NEQIQVAGDCVKEFAMVNVNGVVGAGIDGTTTRITRDEIGTGTGNSLDKSKPHLSKDGINAG 1105  
 Db 397 FGSTDFGNLAVQIVVPNNKI-----LTCNFTGDAKNNGN----- 430  
 QY 1106 GKKITNIQSGEIAKNSHDVATGCKIYDLKTELENKISSTAKTAQNSLHEFSVADEQGNF 1165  
 Db 431 -----TAGVITFNANGTLVSGN-----NIVVTNIKATEVAGIVQLSGIHGAEIRLGNAGSIF 447  
 QY 1166 TVSNPYSSTVTSKTSVITFAGENGITTKVKN--GVYRVGIDQTKGLTTPKLTGVNNN--- 1221  
 Db 448 --TDP-----NIVVTNIKATEVAGIVQLSGIHGAEIRLGNAGSIF 487  
 QY 1222 --GKGVINSQNGQNTITGLSNTLA--NVINDKGSVVRTTEQGN-----IHKDEKTRAAS 1272  
 Db 488 KLADGTVINGPVNQNPVLN--NNALAAGSIQDLSAIIITGIDIGNAVNALQDITLANDAS 546

[illegible]

WAPA OR N17C.  
 Bacillus subtilis.  
 Bacteria; Firmicutes; Bacillus/Clostridium group;  
 Bacillus/Staphylococcus group; Bacillus.  
 NCBI\_TaxID=1423;  
 [1] SEQUENCE FROM N.A.  
 STRAIN=168;  
 MEDLINE=93302506; PubMed=8316082;  
 Foster S.J.;  
 "Molecular analysis of three major wall-associated proteins of  
 Bacillus subtilis 168: evidence for processing of the product of a  
 gene encoding a 258 kDa precursor two-domain ligand-binding  
 protein.";  
 Mol. Microbiol. 8:299-310(1993).  
 [2] SEQUENCE FROM N.A.  
 STRAIN=168 / BGSCIAL;  
 MEDLINE=95219088; PubMed=7704263;  
 Yoshida K.-I., Sano H., Seki S., Fujimura M., Fujita Y.;  
 "Cloning and sequencing of a 29 kb region of the Bacillus subtilis  
 genome containing the hut and wapa loci";  
 Microbiology 141:337-343(1995).  
 -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,  
 MOTILITY, SECRETION OR DIFFERENTIATION.  
 -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED  
 INTO THE MEDIUM.  
 -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS: THE N-TERMINUS, HAS THREE  
 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE  
 C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED  
 MOTIF REPEATED 31 TIMES.  
 -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME  
 SIMILARITY TO THE REPEAT IN E.COLI RHS GROUP OF PROTEINS (RHS-A-D).  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: L05634; AAA22883.1; -  
 EMBL: D31856; BAA06656.1; -  
 EMBL: D29985; BAA06260.1; -  
 EMBL: D83026; BAA11683.1; -  
 EMBL: Z99124; CABI5959.1; -  
 PIR: S32920; S32920.  
 Subtilisin; BG10797; wapa.  
 Pfam: PF02018; CBD\_6; 1.  
 Cell wall; Repeat; Signal.  

SIGNAL	1	28	OR 32 (POTENTIAL).
CHAIN	29	2334	WALL-ASSOCIATED PROTEIN.
DOMAIN	504	869	3 X 101 AA APPROXIMATE TANDEM REPEATS.
REPEAT	504	605	1-1.
REPEAT	636	736	1-2.
REPEAT	769	869	1-3.
DOMAIN	1021	2139	31 X 21 AA APPROXIMATE TANDEM REPEATS OF X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
REPEAT	1021	1040	2-1.
REPEAT	1042	1061	2-2.
REPEAT	1063	1082	2-3.
REPEAT	1083	1102	2-4.
REPEAT	1109	1128	2-5.
REPEAT	1129	1148	2-6.
REPEAT	1150	1169	2-7.
REPEAT	1174	1193	2-8.
REPEAT	1199	1218	2-9.
REPEAT	1219	1238	2-10.
REPEAT	1646	1665	2-11.
REPEAT	1667	1686	2-12.
REPEAT	1690	1709	2-13.
REPEAT	1711	1730	2-14.

FT REPEAT 1732 1751 2-15.  
FT REPEAT 1753 1772 2-16.  
FT REPEAT 1795 1814 2-17.  
FT REPEAT 1820 1839 2-18.  
FT REPEAT 1840 1859 2-19.  
FT REPEAT 1861 1880 2-20.  
FT REPEAT 1887 1906 2-21.  
FT REPEAT 1908 1927 2-22.  
FT REPEAT 1929 1948 2-23.  
FT REPEAT 1969 1982 2-24 (APPROXIMATE).  
FT REPEAT 1983 2002 2-25.  
FT REPEAT 2008 2027 2-26.  
FT REPEAT 2028 2047 2-27.  
FT REPEAT 2051 2070 2-28.  
FT REPEAT 2071 2090 2-29.  
FT REPEAT 2093 2112 2-30.  
FT REPEAT 2120 2139 2-31.  
SQ SEQUENCE 2334 AA; 258329 MW; B75138CCD278BA3 CRC64;

Query Match 4.2%; Score 432.5; DB 1; Length 2334;  
Best Local Similarity 20.3%; Pred. No. 1.9e-08;  
Matches 466; Conservative 272; Mismatches 781; Indels 779; Gaps 118;

QY 182 LKXIQTSDGKIKYRTRAGCHAST-AVGA-MSYAOGHFSNAEGTATAPAAYSLAVGLA 239  
DB 369 LKPIQNMVTKATLKTYVAHSYGTGATGLWLTVNSYDNARVWTKPASKNIG--- 424

QY 240 AQATKSSIAVGSNAKANAFATAIGNTVNLG-----RGVALFGSOTLDRDNT-- 291  
DB 425 -----KADVHKGOWASYDVTAAVKSWNSGGANYGFKLHTNGNGKEYWKLISSANSANKP 479

QY 292 --DASAYVPLGKTLADQYKATROGDSTDFISGNSNNSSIRKLIINVAGSRDRTDAVN 349  
DB 480 YIEVTVTIPIKNT--PTIRAYHNGDSTGYFDISWKKVEGAKYKVIYNGKEYQOALISAGN 537

QY 350 VAQIK-----LVBELANRKITFFGDDGDNNSNSVERGLGNTLTITKGDQATNALTEAN- 400  
DB 538 VTSWSTGKKIWPSTAETASIAKRYKHLHDG---KDGAEALDPSVPYKSGSGSYA-TSKNY 593

QY 401 -IGV---VTDNGKLKVKLAKELTGLTSVSNATKITVSNNTNNAELQSG--GLTFSPITG 454  
DB 594 WIGVSAIFDQEGEAMSAKAPV--IPNVGKAQAPSAKGYNNGNA---TGVFDSLWRAVSG 648

QY 455 TKTDKTVYSIDGLKF---TNDNSNIATKGTTRITRITKKIKGAGTNDGVDESKPYLDNEKL 510  
DB 649 A-TGYKVQVENGKGFETLDLGNQTSWTTKG-----KKI-----WPTSAEI 687

QY 511 KVCNSTLN-----SGS-LTVNNVTGCKQIQVGANGIK-----FATVANN-----VANTSATV 556  
DB 688 KAGYALHLKDGSGAELPIN--PGPTYKNAGDGAKRNYSFKIIAYNKKDGEATASPAATP 745

QY 557 GTARITEEK--IGFAGTNDGVDEQAPYLDKERLKVGRVEITTDGGINAGNHKITGLTNGI 614  
DB 746 ALPDIARPKNVWGTGLTYN-----TKSSQTGVVNLWEKVQNAKGVK-NIYNG- 792

QY 615 ANTDAVITKQKDAKPTLNAGDGISNSNNGDLVSDSGNITPTTYNISVKTTLKNSNGTS 674  
DB 793 -----KEYQ-----SFDVGADHWTTQNKNIWPTSEEIKAGSY-----KLHTDGGK 833

QY 675 GNNKFSVNAHDNNSLVATKDLADYLNKVNETAADSLPSFKVONGDNNNAITVKGDKTNG 734  
DB 834 GELALDPSVYNNAN-----GNYGKKNY-----SFTL-----VAYDANG 868

QY 735 KTFNTLKLK---GENGVNITFNATGTVTGEGIDQSNGLTTPKLTGVSDTNGNRLVIEQVP 791  
DB 869 ETIPTAPFNPTFHEGAFLCTEYEWLSIIDIPSCQLNGAT-----GNVIVNEEDL 917

QY 792 SADGNSTKNIKGLSPTLPSIAS----- 814  
DB 918 SIDGRPG---LGLSRTYNSLSDSDHLFGQGWYADAETSIVISTDGGAMYIDEDATTHFT 974

QY 815 -----PSGRNIALGNTIEE---KDSNAASISDDVLNAGFNLNKNGKDFVSTYDT 862

DB 975 KKADGTYQPETGVYLELTETADQFILKTK-----DOTNAYFNKKGKQLQK----- 1019

QY 863 VDFIDGNATTATVTYDEANOTS-----KVAYDVNVDEKTIETLTDGNGKQKQGVKTIK 914

DB 1020 --VVDGHNNATVTYNDKNQLTAITDASGRKLTFTYDENGCHVTSITGPKNKKVITYSYEND 1077

QY 915 LTE--TSTNGNATFTSDDHIALVKASDIAGNLNTLAEEL-----HTTKGTANTALQ 964

DB 1078 LLKKVTDTDGTVTSYDYDSEGRLVK--QYSAN--STEAKPVFTEYQYSGHRLERAINAKKE 1134

QY 965 TFTVKKVDENDKA-----DDTNAITVCKDGTSGKVNT--LKLKKNGL-- 1005

DB 1135 TY--VYSYDADKKTLLMTQPNGRKVQYGYNEAGNP IQVIDDAEGLKITNTKYEGNNVVED 1193

QY 1006 ----DKT-----DKDGTVT-----EGINTQSLKAGDSTTLNNGLSIKNTA--- 1044

DB 1194 VDPNDVGTGKATESYQYDKDGNVTSVKDAYGTETIYFNKNDVTKMKDTEGNTVDIADYG 1253

QY 1045 ----SNEQIOVGADGVKFAMVNNGVGA-GIDGTTTRIRD---EIGFTGTNGSLDKSKP 1095

DB 1254 LDVSETDQSGKSSAAVYKYGNOIQSSKDLASTNILKDGSPQAOKSGWNLTASKDRR 1313

QY 1096 HLS----KGINAGGKKI-----TNIOSGEI----AK 1119

DB 1314 KISVIADKSGVLGSKALEVLSQSTSAGTDHGYSSATQTVLEPNTTYTTLGKIKTDLAK 1373

QY 1120 -----NSHDVGT----- 1127

DB 1374 SRAYFNIDLRDQKRIQWIHNEYSALAGKNDWTKROITFTTPANAGKAVVMEVDHDK 1433

QY 1128 ---GKIVDLKTELE-NKISSTAKTAQNSLHEFSVADQGNNTVSNPYSSYDTSKTS DVI 1183

DB 1434 DGKGAWFDEVOLEKEGVSSSNPVQNS--SFTSATE---MNVSG--ASVDS----- 1479

QY 1184 TPAENGCIITTKVNGVVRVIGDQTKGLTTPKLTGVGNN-NGKGIIVNSQNGQNTITGLS-- 1240

DB 1480 ---EEGFNDVSLKAARTSASQAGSVTKOTVVLGOSANDKPVYL-----TLTGMSKA 1528

QY 1241 -----NTLANVTNDKGSV-----RTTEQGNIIKDEKTRAASIVDVLVSAG 1280

DB 1529 SSVKFTDEKDYSLQANVTYADGSTGIYNAKPPSGTQWNNRAAVPIPKTPINKVDI-SIL 1587

QY 1281 FNLQNG-----EAVDFV-----STYDVTNFANGNTT-----AKVYDDT-SKT 1319

DB 1588 FOKSATGVWFDDIRLIEGSLITKSTYD---SNGNVVTKREDEELGYATSTDYDETCKKT 1643

QY 1320 SKVYDVNVDDTII-----EVKDKKLCVKVKTTLTSTGTGANKFALSNOATGD----- 1366

DB 1644 SET--DAKGEKTTVYTDQADQLTNMTLSNGTSLHSDYKDEGNEVSKTIRAGADQTYKFEY 1701

QY 1367 ---ALVKASDIVAHLNLTLSGDIQTAKGASANNAGSYVDADGNKV--IYDSTD---NKY 1417

DB 1702 DVMGKLVKTTDPLG--NVLASEYDA-----NSMLTTIISPNGNEVSLSDGTDTRVSKS 1753

QY 1418 YQA--KNDGTVDKTEKAVAKDLVAQAQTPDGLAQMNVKSVINKEQVNDANKKOGINEDN 1475

DB 1754 YNGTEKYIFTYDKNG-----NETSVVNEQ--NTTKKRTEDKN 1790

QY 1476 AFVKGLEKAAADNKTNAAVTVGDLNVAQOTPLTFAGDTGTAKKLGLETLIKGGQTDTN 1535

DB 1791 RLTELTDGRGS-----QT-WTYPSP-----SDKLKTFSIHGDQGTN 1827

QY 1536 KLTDDNNGVYVAGTDGFTVKLAKDLTLNLSVNAAGTKIDEKG--ISFVDANGQAKANTPVL 1593

DB 1828 QFTYN-----KLDQMIEMKDDSTSSYSFYDENGWVQFTITGNG----- 1865

QY 1594 SANGDLGGKVISVNGKGTKTDAAANYQQLNEVRNL---LGLGNDNADGNQVNIADIKD 1650

DB 1866 -----GGTSFS-----YDERNLVSSLHIGDKN--GGDILTESYEYD 1899

QY 1651 PMSGSSSNRTVIK---ACTVL--GGKGNNDTEKLTATCGVQGVGDKGNANGDLSNWWVKT 1705



Ddb		1900 AN----	GNRTTINSASGKVQYEGYKLNQLNFKETHEDGTVIEWTYDFGN-----RKTVT 1951
Qy		1706 QKDGSKKALLATNAAGOTNTVNNPAEADRINEOGI-----RFFHYND--- 1750	
Ddb		1952 IKDGSKTVNASFNIMNQ-----LTKVNDESISYDKNGNRTSDGKFYTWDAD 2000	
Qy		1751 -----GNOEPVV-----QGRNGIDSSASGK-----HSVATGFQAKDAEAAVAI 1789	
Ddb		2001 NLTAVTKKGEDPFAATKYDEKNRIQKTGVNKVTNYFYDGDSLNVLYETDANNV---- 2056	
Qy		1790 GROTQAGNSOIAIGNAO-----ATGD-QSIAITGTGNVAVAGKHSGAIGD 1832	
Ddb		2057 -----TKSYTYGDSGQLLSYSTENGKKFYFHYNAHGDIITAISDSTGKTVMKYQYDAWN 2109	
Qy		1833 PSTVKA-----DNSYSVGNNGNOTDAT-----QTDFV-----GVGNNTIVTES 1870	
Ddb		2110 PTKTEASDEVKDNRYIA-GYOYDEETGLYYLMARYYEPRNGVFSLDDPPDGGSGDSDLQ 2168	
Qy		1871 NSVALGNS-----ALSAG--THAGTOAKSKSDGTAGTTTTAGATGTVKGFA 1914	
Ddb		2169 NGAYGNPNVMDPDGHVWLVYNAGFAAYDGYKVGSG-----KGWK 2213	
Qy		1915 GOTAVGAVSGASGABERRIONVAAGEVSATSTDVANGSQLYKATQSIAIANATNELDHRIHQ 1974	
Ddb		2214 GAAWAAA-----SNFGPKIKF-----GASRAYKFTKKAVKITGHTRHGLNQ 2255	
Qy		1975 N-----ENKANAGISSAMAMASMPQA---YIPGRSMVTGCIATHNGQGAVAVGL 2020	
Ddb		2256 SIGRNGRGVNLRAKLINA-VRSPKKVIKOPNCATKYVCKKATV---VLNKRKGVITAYGS 2311	
Qy		2021 SKLSDNGQWYFKINGSAD 2038	
Ddb		2312 SR-AKGSKHVEHTHGKGN 2328	

RESULT 10  
PIP\_LACLC STANDARD;      PRT;    1902 AA.

AC	PI6271;
DT	01-APR-1990 (Rel. 14, Created)
DT	01-APR-1990 (Rel. 14, Last sequence update)
DT	01-NOV-1995 (Rel. 32, Last annotation update)
DE	PI-TYPE PROTEINASE PRECURSOR (EC 3.4.21.-) (WALL-ASSOCIATED SERINE PROTEINASE).
GN	PrpP.
OS	Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG	Plasmid pMW05.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC	Lactococcus.
OX	NCBI_TaxID=1359;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-WG2;
RX	MEDLINE=86149035; PubMed=3278687;
RA	Kok J., Leenhouts K.J., Haandrikman A.J., Ledebor A.M., Venema G.;
RT	"Nucleotide sequence of the cell wall proteinase gene of
RT	Streptococcus cremoris Wg2.";
RL	Appl. Environ. Microbiol. 54:231-238(1988).
CC	-!- FUNCTION: PROPEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC	GROWTH OF THE BACTERIA ON MILK.
CC	-!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC	SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
CC	E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC	PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS.
CC	ALTHOUGH IT HAS BEEN SHOWN TO HYDROLISE HEMOGLOBIN AND OXIDIZED
CC	INSULIN B-CHAIN.
CC	-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC	SUBTILASE FAMILY.

---

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL; M24767; AAAL767.1; -  
CC HSSP; Q99405; IMPT.  
CC MEROPS; S08.019; -  
CC InterPro; IPR000209; -  
CC InterPro; IPR001899; -  
CC Pfam; PF00746; Gram\_pos\_anchor; 1.  
CC Pfam; PF00082; Peptidase\_S8; 3.  
CC PRINTS; PF00723; SUBTILISIN  
CC PROSITE; PS00136; SUBTILASE\_ASP; 1.  
CC PROSITE; PS00137; SUBTILASE\_HIS; 1.  
CC PROSITE; PS00138; SUBTILASE\_SER; 1.  
CC PROSITE; PS00343; GRAM\_POS\_ANCHORING; 1.  
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;  
Transmembrane.  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT ACT\_SITE 1867 1872  
FT DOMAIN 1867 1872  
FT FT  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620

QY 478 --TKGTRITKK-KIGFAGTNDGVDESKPYLDNEKLVGNSTLNSGSLTVNNTGNKQIQ 534  
Db 483 ADAGKIAIVKRGELSF-----DDKQYAAAGAAAGLIIVNN-----519  
QY 535 VGANGKIFATVANNVANTSATVGTARITBEKI-----GFAGTN 572  
Db 520 ---DGTATPVTSMALITTPFTFGLSSVTGQKLVDMWTAHPDDSLGVKIALTLVPNQKYTE 576  
QY 573 DGVEQAPYLDKRLKVRVEITDGSINAGNHKITGLNGIANIANTDAVITKOLKDAKPTL 632  
Db 577 DKMSDFTSY-----GPVSNLSFKPDITAPGNGIWTQNNNGYTNMSGTSM-----ASPEI 626  
QY 633 NACDGI---SINSGNDLVDSNNITPTYNISVKTKLNS-----NCTSGNNKFSVSNH 685  
Db 627 AGSQALLKQALNNKPNFYAYYKQLKGTALTDFLKTEMNTAOPINDINYNVI-VSPRR 685  
QY 686 DNNSLVTAADLADYLNKVNET---ADSALEPSFKVQNGDNSNNAITVGKDTNGKTFNTLKLK 743  
Db 686 QGAGLVDDVRAAIDALEKNPSTVVAENGYPAVELKDFSTSD-----KTF---KL- 730  
QY 744 GENGVNITNTRATGTVTFGIDQSNGLITPKLVGSDTNGRLVIEQVPSADGNSTNIIK 803  
Db 731 ---TFTNSTTHELTQMD-SNTDTNAVYTSATDPSNGVLKYDKKIDGA-----AIK 776  
QY 804 GLSPTLPSTASPSGRNIALGNTIE-EKDKSNAASIDDVLN-----AGFNLK-----848  
Db 777 AGS-----NITVPAGKTAQIEFTLSLPSFKDQOQFVEGFLNFKSGDSRLNLPYMGFFGDW 832  
QY 849 NNGKDKDFVS-----TYDVTDFIDGNATAT-----VTYDEANQTSKVAIDVNVND 893  
Db 833 NDCKIVDSLNGITYSPAGNFGTVPLLT-NKNTGTOYGGWVTDAQGNQI-----VD 883  
QY 894 EKTIELTGDNKG-KOLGVKTIKLTETSTNGNATTFSTDDHALVKASDIAGN-LNLTAE 950  
Db 884 DOAIAFSSDKNALYNDISMKYLLRNIS-----NVQVDILDQOGKNKVTLLSS 930  
QY 951 EIH7TK-----GTANTALQTFVKKVDE-NKAD- 978  
Db 931 STNLTKTYNNHAHQYIYNAPAWDGTYYDQRDGNKLTADDGSYTYRISGVEGDKRQV 990  
QY 979 -----DTNAITVGKDGTSKGVNTLKLKNGKIDIKTKDG-----TVTFGINTOSGLK 1026  
Db 991 FDVPFKLDSKAPVHRVALSAKTENGKTOYLLTAEKADLGLDASKVKTAINEVNLD 1050  
QY 1027 A--GDSTTLNNGLSKNTASNEQIOVGAGVKFAMVNGNNGVAGIDGTTTRITRDEIGFT 1084  
Db 1051 ATFTDAGTTADGYTKIETPLSDEQAQ-----LGNG-----DNSAELYLTDNASN 1095  
QY 1085 GTNGLDKSKPHLSKDG--INAGG--KKITNIOGSETAKNSHDAVT-GGKIYDLKTELEN 1139  
Db 1096 ATQDDASVQKPGSTSFDLIVNGGIPDKISSTTTG-----YEANTQGGGTYYTFSCTYPA 1149  
QY 1140 KISSTAKTAQNSLHEFSVA-DEQGNFTVSNPNYSYSDTKSDTSDVITFAGENGIT---TTK 1194  
Db 1150 AVDGTYNAGKKHDLNTTYDAATNSFTASMPVTNADYAAQVDLYADKAHTQLLKHFDTK 1209  
QY 1195 VNGGVVRVIGIDQTKGLTTPKLT-VGNNGNG-----GIV-----INSQNGQNTITGL 1239  
Db 1210 VR-----LTAPTFTDLKFNNGSDQTSSEATIKVTGTVSADTKTVNVGDTVAAL 1256  
QY 1240 -----SNTLANVTNKGSVTTEOGNIKDEKTRAASIVDVLISAGFNQ 1284  
Db 1257 DAQHHSFSDVPVNVGNTIKVATDEGNTTTEOKTITSSYDP-----DMLK-----1303  
QY 1285 GNGEAVDFSVTYDVFANFNGNTTAKVYDDTSKTSKVYDVNVVDDTTIEBKKLGVK- 1343  
Db 1304 -NSVTFDQGVTFGANEF---NATSAKF-YDPKGTGIATITGKVKHPTTLQVQDKQIPDK 1358  
QY 1344 -----TTTLTSTGTGANKFAL-----SNOATGDALVKASDIVAHLNLTSLDQITAKGASQ 1393  
Db 1359 DLTFSTLGLTGLQKPFVGVVGGDTTQNKTFQEQALTFILDAVA--PTLSLESST-----1410  
QY 1394 ANNSAGYVDADGNKVIYDSTDN-KYYQAKNDGT-----VDKTEKVAOKLVAAQAPDGD 1446

Db 1411 --DAPVYNDPNFOITGTATDNAAOYLSLSINGSSVASQYVDININSKPGHMAIDQ-PVK 1467  
QY 1447 TLAQMNVKSVINKEQVNDANKKOCINEDNAFVKGLEKAAASDNKTNAAVTVGDLNAVAQT 1506  
Db 1468 LLEGKNVLTV-----AVTD-----SEDNTTKNITVYEPKTL-AAPTV-----T 1507  
QY 1507 PLTFAGDTGTTAKKLGTELPIKGGQDTNTKLTNNIGVWAGTGTGTVKLAKDLNLNSVN 1566  
Db 1508 P-----STTEPAKTVTLTANSAAT-----GETVOYSAD-----1535  
QY 1567 AGG---TKIDEKGISFYDANGQAOKANTPVLSANGCLDGGKVISNVGKTRDQDAANYQOL 1623  
Db 1536 -GGKTYODVPAAGVT-VTANGTFKFKSTDLGYNESPAVDYVVTNI-----KADDPAQLOAA 1589  
QY 1624 -NEVRNLLGLGNDNADGNQVNIADIKKDPNSGSSNRTVIKAGTVLGGKGNNDPEKLTATG 1682  
Db 1590 KQELTNL-----IASAKTSLASGKYDDATT---TALAAA---TQKAQTA 1627  
QY 1683 --GVQGVVDKGNANGDLNWNVKTQDKGKKALLATYNAAGQPNYVYNNPAAIDRINE 1740  
Db 1628 LDOTNASVDSLITGANRDLQT-----AINQLAAKLPAKKTSL-----1665  
QY 1741 QGIRFFHVNDGNOEPVVOGRNGIDSSASGKHSVAIGFOAKADGEEAIVAIGROTQAGNOSI 1800  
Db 1666 -----NOLQSVKAALGTDGLNQTDPSTGKTFTAAALDDLVA-----QAQAGTQT- 1708  
QY 1801 AIGDNAQATGDQSTAICTGTVNVAGKHSAGIDPSTVKADNSYSVGNNNQFTDATQTDVFG 1860  
Db 1709 --DDHQAT---LAKVLDVAVLAKLAEG-----IKAAATPAEYGNK---DAATGKTWY 1752  
QY 1861 VGNNTVTVESNVALGNS-----SAISAGTHAGTQAKKSDGTAGTATTTAGATG 1908  
Db 1753 ADIADTLTSGOASADADSLAHLQALQSLKTKVAAAVEAAKTVGKGDTTSDKGGGQ 1812  
QY 1909 TVKGFAGQTAGVAVSVCASGAERIQNVAAGEV-----SATSTD 1947  
Db 1813 TPAPAPG-----DIGKDKGDEGQSPSSGNGIPTNPATTSTSTD 1851

RESULT 11  
P3P\_LACLC STANDARD; PRT: 1902 AA.  
ID P3P\_LACLC AC PI5292;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE PLII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-  
DE ASSOCIATED SERINE PROTEINASE).  
GN PRTP.  
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
OG Plasmid.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Lactococcus  
OX NCBI\_TaxID=1359;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.  
RC STRAIN=SK11;  
RA MEDLINE=89340435; PubMed=2760036;  
RX Vos P., Simons G., Siezen R.J., de Vos W.M.;  
RT "Primary structure and organization of the gene for a procaryotic,  
RT cell envelope-located serine proteinase";  
RL J. Biol. Chem. 264:13579-13585(1989).  
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
CC -!- GROWTH OF THE BACTERIA ON MILK.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
CC SUBTILASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; J04962; AAA03533.1; ALT\_S80.

PIR; A32634; A32634.

HSSP; P00782; 2S8T.

MEROPS; S08.019; -.

InterPro; IPR000209; -.

InterPro; IPR001899; -.

Pfam; PF00746; Gram\_pos\_anchor; 1.

Pfam; PF00882; Peptidase\_S8; 3.

PRINTS; PR00723; SUBTILISIN.

PRINTS; PS00136; SUBTILASE ASP; 1.

PROSITE; PS00137; SUBTILASE HIS; 1.

PROSITE; PS00138; SUBTILASE SER; 1.

PROSITE; PS00343; GRAM\_POS\_ANCHORING; 1.

Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;

Transmembrane.

SIGNAL 1 33

PROPEP 34 187

CHAIN 188 1902

DOMAIN 188 1876

TRANSMEM 1877 1895

DOMAIN 1896 1902

ACT\_SITE 217 217

ACT\_SITE 281 281

ACT\_SITE 620 620

DOMAIN 1867 1872

SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;

Query Match 4.1%; Score 423.5; DB 1; Length 1902;

Best Local Similarity 20.4%; Pred. No. 3.le-08;

Matches 458; Conservative 230; Mismatches 817; Indels 735; Gaps 104;

20 AYAASHTGGGSCATQGVSVTLSPARALAVLIGATLNGSAYAAQIITK-IEIQQ 78

31 AKAAISOOTKGSSLA-----NTVTAATAKQAATDTTAATTN-QAIALQAAGIDYKN 82

79 TNKI-----NNTLKGDALEGAETAFSLSKAQSQAIAIGSYK 118

83 LNKVQQDDIYVDVIVQMSAASENGILRTDYSIAEQOETNKVIAAGAAVAAEQVT 142

119 PDPNNGSGNV-----GSHAKGNESTAI-----141

143 QOTAGESYGYVNGFSTKVRVVDIPKLQIAGYKTVTLAKVYPTDAKANSMAVQAVWS 202

142 -----GGDLVLAEGDASTAIGSDLLYLPKNLDLK-----NEFKLI-HGHEILKKI--- 185

203 NYKKGECTVSVVIDSGIDPDKHMLSDDDKDKLTKSDVEKFTDVKHGRYFNSKVPYG 262

186 -----OTSDGKKYRRTAQAQHASAVGAMSYAAGHESNAGFTYATAEAAVSLAVGL 238

263 FNYADNNTITDDKYD-----EQHGM-----HVAGIIGANGTGDPAKSVGV 305

239 AAAA-----TKQSSIVAGSNAKANAFATAIGGTVVNLGRGVVALGFGSQILDRDN 289

306 APEAQLLAMKVFNSDTSKTSATVWSAIEDSAKIGADVLNMSLG--SNSGNOTLE--- 360

290 NTDAASYPLGLTLADQYKATPQGSTDIFSGTSGNNNSIRRKIIIVGAGSRDITAVN 349

361 -----DPELAAVQANESGTAA-VISAGNSGTSGSA-----TEGVN 395

350 VAQLKLVEELANRKTIFKGDGN-----NNSVERGLNGTLTKGDAQTNALTEANIGVVD 406

396 K-----DYGLQDNEMVSGPSGTSG--ATTVASAENDVITQA--VTITD 436

407 GNLGKVKLAKELTGLTSVSAATNKITVNTNNAELQSGLTFSPITGKTDKTVYSIDG 466

437 GTGL--QLGPETIQLSS-----HDFGSGFDOKREYIV-- 466

QY 467 LKFTNDSNSIATKGT-----TRITKKKIGFAGTND-GVDESKPYLDNEKLVKGVNSTLNGS 521

DB 467 -----KDSGNLSKALADYTADAKGKIALVKRGESFDDKOKYA-----QAAGAAG 513

QY 522 LRVNNTTGNKQIQVANGIKPATVANNVANTSA-TVCTAHITEKI-----GPAQTNDG 574

DB 514 LLIIVNTDGT-----ATPMTSIALTTTFTFGLSSVTKGKLVWVTAIPDDSLG 561

QY 575 VDEQAPYLDKERL-----KVGREVTITDGSINAGNHKITGLTINGIANTDAVTI 622

DB 562 VKITLAMLPNQKYTEDKMSDFTSYGPVSNLSFKPDITAPGGNINWSTQNNNGYTNMSGTSM 621

QY 623 KOLDAKPTLNAGDGI-----SINSNNGDLVDSSNGTITTYINISVKTTLNS-----NGTSG 675

DB 622 -----ASPFIAQSQAALLKQALNNKNPYYAYKQLKGTALTDLKATVEMTAQPTNDINY 676

QY 676 NKFYSVNAHDNNSLVTAADLADYLNKVNET--ADSALPSPKVVQNGDNNNAITVCKDTN 733

DB 677 NNVI-VSPRRQAGLVDRKAAIDALEKNFSTVVAENGYPVELKDFSTSD-----725

QY 734 GKTFTNLKLGKENGVNITTNRAITGVTFEGIDOSNGLTTPKLTIVGSDTNGNRLVIEQVPSA 793

DB 726 -KTF-----TFTNRTHETLYQMD-SNTDTNAVYTSATDPNSGVLYDKRIDGA 773

QY 794 DGNSTNIIKGLSPTLPSIASPSGRNIALGNTIE-EKDKSNAASIDDLN-----AGEN 846

DB 774 -----AIKAGS-----NITVPAGKTAQIEFTLSLPSFDOQFVSGFLNEKSGDSRLN 822

QY 847 LK-----NNGKDKDFVSTYDVFIDGNATVATVYDEANQTSKVAYDVNVDEKTTIE 898

DB 823 LPYMGFFGDNDCK-----IVDSLNG-----ITYSPAG-----850

QY 899 LTQDNG-----KKQGVKTIKLTETSTNGNATTFSTDDHDLVKAASDIAGNLNTLAEI 952

DB 851 --GNFGTVPLLNKNTQYVGMVTDADGNKV-----DQAIATFSSDKNALYNDISKY 904

QY 953 HTTGTGTALOTFTVKKVDENDKADDTNATTVGKDGTSKGVNTLK-----998

DB 905 YLLRNISNVQDIL-----DQGNKVTTLSSSTNRKKTYYNAHS 943

QY 999 -----LKGKGLDITDKDGTVTFTGINTOSGLKAG-----1028

DB 944 QOYIYVNAWPGTYDQDQDQ-NIKTADGSIYTRI---SGVPEGGKRQVDFVPFKLDS 999

QY 1029 DSTTLNNGSLTKN-----TASNEQIQVAGDGVKVFAMVNVGV-----CAG 1069

DB 1000 RAPTVRHVALSAKTENGKTQYLLTAAEKDDLGLDQATKSVKTEINEVTNLDATFDAGTT 1059

QY 1070 IDGTTRI-----TRDEIGFTGTNGSLDKSKPHLSKDG-----INAGG 1106

DB 1060 ADGYTKIETPLSDEQAQALGNSDASAELYLDNANATDQDASVQKPGSTSFDLIVNGCG 1119

QY 1107 --KLTNIOGSEIAKNSHDAVT--GGKIYDLKTELENKISSATAQNSLHFEFSA-DEQG 1162

DB 1120 IPDKISSITIG-----YEANTQGGTYTFSCTYPAAVDGTVDQAQKKHDLNTYDAAT 1173

QY 1163 NNFTVSNPSSYDTSKTSVITFAGENGIT-----TTKVNKGVVRVICIDQTKGTTPKLT-V 1217

DB 1174 NSFTASMPVTNADYAAQVDLYADKAHTQLKLFHFDKVR-----LMAPTFTDL 1220

QY 1218 GNNNGK-----GIV---INSONQNTITGL-----SNTLIANTYN 1248

DB 1221 KFNNGSDQTSSEATIKVTGIVSADTKTVNVGHVVAALDAQHFEVDVVPVNYGONTIKVIAT 1280

QY 1249 DGKSVRTTEOGNITKEDKTRAASIVDVLISAGENLOGCEAVDFVSTYDT-VNEFANG--N 1305

DB 1281 DKDGNTEQKTISSYDP-----DMLKKS-----TFDQGVKFGTNRKN 1320

QY 1306 TTTAKVYDDTSKTSKVVDVNVDDTTIEVKKKLGVK-----TTTLTSTGTCAKFKAL- 1359

DB 1321 ATSAKF-YDPKGTIATITGVKKHPTTLQVDGKQIPIKDDLTFSFTLDLTGCRPFGVV 1379

QY 1360 -----SNOATGDALVKASDIVAHLNLTLSGDIQTAKGASQANNSAGYVDAGNKVIYDSTD 1414

```
Db 1380 VGDTONKTFQALSFILDAVA--PTLSUDST-----DAPVYNDPNFQITGTATD 1429
Qy 1415 NKYYQAKNDGTVDKTEKAKDLVAQAQPDGTLAQMNVKSVINK-EQVNDANKKQG---- 1470
Db 1430 NAQYLS-----LSINGSSVASQVEDININSKPGHMA 1461
Qy 1471 INEDNAFYKG---LEKAASDNKTKNAVTVGDLNVAQTPPLTFAGDTGTAKKLETIT 1527
Db 1462 IDQPKLLGKKNLVAVVTDSDN---TTTKNITVYEPKKTAAFTVPS-----TT 1511
Qy 1528 KGGQDTNKLTDNNIGVWAGTGTGFKLAKDLTNLSNVNAGG---TKIDEKGISFVDANG 1584
Db 1512 EPAQIVT--LTAN-----AAATGEIVQISAD-----GKTYQDVPAAGVT--ITANG 1554
Qy 1585 QAKANTPVLSANGLDLGGKVIISNVGKTKDTPDAANVQQL-NEVRNLLGLGNADGNNOVN 1643
Db 1555 TFKFKSTDLYGNEPAVDYVVTNI-----KADDPQAQLQAQKQELTNL----- 1596
Qy 1644 IADIKDPNPGSSSNETVFKAGTVLGGKGNNDTEKATG--GVQGVGDKDGNANGDLSNV 1701
Db 1597 IASAKTSLASGKYDDATT-----TALAAA-----TQKAQTALDQTNASVDSLITGANRDLQT- 1647
Qy 1702 WKTKQDKGSKKALLATYNAGQNTVYNNPAAEIDRINEQIRFHVNDGNOEPVVOGRN 1761
Db 1648 -----AINQLAALPADKKTSL-----NQLQSVKDAL 1675
Qy 1762 GIDSSAGSKHSAVIGFQAKADGAAVAIGROTQAGNQSIAIGDQAQTDGQSTAGTGNV 1821
Db 1676 GTDLGNQDTPSGKTFATLDDIVA-----QAQAGTQT--DDQLQAT--LAKILDEV 1723
Qy 1822 VAKKHSAGLGDSTPVKADNSYSGVNNQTFDATQDVFVGNNITVTSNSVALGNS---- 1878
Db 1724 LAKLABG-----IKAATPAEYGNAK-----DAATGKTWYADIADTLTSCQASADSKLA 1773
Qy 1879 -----SAISAGTHAGTQAKKSDGTAGTTTAGATGTVKGFAGQTAGVAVSGV--A 1926
Db 1774 HQAQLSKTKVAANAVEAAKTVKGDDTGTGSDKGGGQGTAPAPAGDTGKDKGDEGQPS 1833
Qy 1927 SGAERRIONVAAGEVSATST 1946
Db 1834 SGG-----NIPKPTATTST 1848

RESULT 12
CBPA.CLOCL
ID CBPA.CLOCL STANDARD; PRT; 1848 AA.
AC P38058;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CELLULOSE BINDING PROTEIN A PRECURSOR.
GN CBPA.
OS Clostridium cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1493;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92228810; PubMed=1565642;
RA Shoseyov O., Takagi M., Goldstein M.A., Doi R.H.;
RT "Primary sequence analysis of Clostridium cellulovorans cellulose
binding protein A.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3483-3487(1992).
CC -!- FUNCTION: BINDS TO CELLULOSE FIBERS AND COORDINATES CELLULOSE
ENZYMES.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- PTM: CONTAINS ABOUT 10% OF CARBOHYDRATES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M73817; AAA23218.1; -.
DR PIR; A44140; A44140.
DR HSP; Q06851; INBC.
DR InterPro; IPR001956; -.
DR InterPro; IPR002102; -.
DR Pfam; PF00942; CBD_3; 1.
DR Pfam; PF00963; Cohesin; 9.
KW Cellulose degradation; Cell wall; Repeat; Signal; Glycoprotein.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 1848 CELLULOSE BINDING PROTEIN A.
FT DOMAIN 29 189 CELLULOSE-BINDING (BY SIMILARITY).
FT SEQUENCE 1848 AA; 189152 MW; 85FA6CE6F771AF1A CRC64;
SQ
Query Match 4.0%; Score 410.5; DB 1; Length 1848;
Best Local Similarity 21.3%; Pred. No. 8.e-08;
Matches 443; Conservative 252; Mismatches 786; Indels 603; Gaps 103;
Qy 181 ILKKIKTS-----TDGKIKYRRT--AOG-----HASTAVGAMSVAQGHFNAFGTYA 226
Db 52 IIKITWSDSLNLDNDVKVRYTSDGTQGTQWCDHAGALLG-----NSYVDNT 101
Qy 227 TAAEAYSLAVLAAQAATKQSSIAVGSNAKANAPA---ATAIGNTVYVNLGRGVALFGSQ 283
Db 102 SKVTANFVKETASPTSYTYVEFG-----FASGRATLKKGFITIOGRITKSDWSNY 154
Qy 284 ILDRDNTTASAVPLGKTLADQY---KATROGSDTDIFSIGNSNNSNNSIRKKNV 339
Db 155 TQTNDYSFSDASSTPVYVNPVKVTGYIGGAKVLGTAPGDPVS-----SIINPT 201
Qy 340 AGSRDQDVAENVVQALKEELANRKITPKGDGNNNSVERGL-----GNLTITK----- 388
Db 202 SATFDKNTVKQADVKTTMTLNGN--TFKITTDANGTALNASTDYSGNDVTSKAYLAK 259
Qy 389 -----GDAQTNALT-----BANIGVYTDGNGLKVKYKLAKELT-----G 420
Db 260 QSVGTTTLNPNFSAGNPQKLVITVVDTPVEAVTATIGKQVNAVAGTAVVPLNTKVPAA 319
Qy 421 LTVSVATNKTIVSTNNNAELQSGGLTSP---ITGKTDKTVYSI----- 464
Db 320 LATIEL--PLTFDSASLEVYSITAGDIVLNPVSFSTVSGSTIKLLFLDDTLGSLQITK 377
Qy 465 DGLKETNDSNSIATKGT--RITKKKIGFAGTNDGVDKPYLDNEKLKVGNSLNSGSLT 523
Db 378 DGVFAITTFKAKAITGTTAKVTSVKL--AGT-----PVGDAQQLQEKPCAVNPGT 427
Qy 524 VNNTTGNKQIQVGANGIKFATVANNVANTSATGTARITEEKIGFAGTNDGVDQEQAPYLD 583
Db 428 INFIDNRMQISVGTATVK-----AGEIAAVPVTLTSPST---GIATAEAQVSEFDATLLE 479
Qy 584 KERLKYRVEITTDSCINAGNHKTLGTNGIANTDAVTIKLKADKPT--LNAGDGSINS 642
Db 480 VASVTAG--DIVLNPVTN-----FSYTVNGNVIKLLFLDDTLGSLISKDGVFVTI 528
Qy 643 N-NGDLVDSGNTTPTTYNISVKTTLKNSNCTSG--NNKFSVSNNAHDNNSLVTAKLADY 699
Db 529 NFKAKAVTS--VTTTP---VTVSGTPVFADGTLAEVQSKTAAGSVTIN---IGDPLEPT 580
Qy 700 LNKVNETADSALEPFRVQNGDNNNAITVGKDTNGKTFNTFLKLGNGVNTNTRATG-- 757
Db 581 ISPVATFEDKKAPA-----DVATMTLNGYTFN-----GITGLTSDYSISGNV 624
Qy 758 -----TVTFGIDQSNGLTTPKL-----TVGSDT--NGNRLVI 787
Db 625 VKTSQAYLAKQPVGDLTLTFNFSNGNKATATAKLWVSIKDAKPTVTTATVGTATVAGETVA 684
Qy 788 EQV--PSADGNSYKNIKGLSPTLPSASPGRNIALGNTLIEEKDKSNAASI-----DDV 840
Db 685 VPVTLNSVSGISAEQLQSFADTLLEVVSVITAGDIVLNPVSFSSVNGSTIKLLFLDDT 744
```

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

1760 RINGDSSASGKHSVAIG---FOAKADGEAAVAIGRQTQAGNQIAIGDNRQAQTGDOSIAI 1816  
 1653 -KCTDYTTISGS-TVTISKAVLATLADGSATL-----BEVFNQGSASAKRLUTIV- 1698  
 1817 GTGNVAVAGKHSALGDP-----STVKADNSVSVGNNOFTDATOTDVGNGNITVTESNS 1872  
 1699 -----PAVVDPPVTDFAVKIDKVSANAGSTVKVPVSLINYSKVG-NVCVAE-YK 1745  
 1873 VALGNSAISAGTHAGTQAKKS-----DGTAGT---TTTACATGTVKVG 1912  
 1746 ISFSSVLTYTGTAGTSIKNPVNFSSQLNGNTITLLFFDNTIGNELLITADGQFATI-- 1803  
 1913 FAGQFAGVAVSGVAGSABERRIQNVA---AGEVATSTDAVNGS 1952  
 1804 ---EFKNAATSGTTAEVKVATISSFADASLTETITKVATVNGS 1844

RESULT 13

HLIA\_SERMA STANDARD: PRT: 1608 AA.  
 ID HLIA\_SERMA AC P15320;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-NOV-1990 (Rel. 16, Last annotation update)  
 DE HEMOLYSIN PRECURSOR.  
 GN SHLA.  
 OS Serratia marcescens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Serratia.  
 OX NCBI\_TaxID=615;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.  
 RC STRAIN=SN8;  
 RX MEDLINE=88257037; PubMed=3290200;  
 RA Poole K., Schiebel E., Braun V.;  
 RT "Molecular characterization of the hemolysin determinant of Serratia marcescens";  
 RL J. Bacteriol. 170:3177-3188(1988).  
 CC -|- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY DEFINED.  
 CC -|- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA REQUIRES SHLB FUNCTION.  
 CC -|- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
 CC -|- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC EMBL; M22618; AAA50323.1; --  
 DR PIR; A28182; A28182.  
 KW Hemolysis; Toxin; Outer membrane; Signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 1608 HEMOLYSIN.  
 FT SEQUENCE 1608 AA; 165078 MW; D669B476FE7DAD51 CRC64;

Query Match 3.9%; Score 404; DB 1; Length 1608;  
 Best Local Similarity 20.1%; Pred. No. 1.3e-07;  
 Matches 398; Conservative 255; Mismatches 702; Indels 624; Gaps 97;

QY 219 SNAGFYATAEAYSILAVGLAAQAATQKQSIAGVSNKANAFANTAIGTIVNVL---GR 274  
 Db 3 NNNFRLSAGKLAALAILLAASAGAYAAEIVAANGANGPGVSTAATGAQVVDIVAPNGN 62  
 QY 275 GVA-----LFGSOILDRDNNNTDASAVVPLGKTL- 303

841 LNAFNLNK-----NGKDKDFSTYDITVDIFDGNATTATVTVDEANQTSKVAYDVNVE 894  
 Db 745 LGSOLISKDGVFATINFAKASVSTVTPVKVSGTVPFADGTLAE-----LSYETVAGS 798  
 QY 895 KTIETLGDNGKQKQGVKTIKLTETSTNGNATTTSTDDHAKVKASDIAGNLN----- 946  
 Db 799 VTINAIGP-----VKTIV-----TATVGTATVKSGETVAVPVLTVSNVPGIATAEQLQSF 847  
 QY 947 -TLAEIHTTNG-----TANTALOTETVKKVDENDKADDT-NALTVGKDGTSKVNVT 996  
 Db 848 ATLEVASITVGDIVLNPVNFSSVNGSTIKLL-----FLDDTLGSLISKDGV---LAT 900  
 QY 997 LKLGKNGLDIKTKD---DGTVTETGINTQSGLKA-----GDST 1031  
 Db 901 INFKAQTVTSTVTPVAVSGTVPFADGTLAELOSKTVAGSVTIEPSPQVKVTVATVGTAT 960  
 QY 1032 TLNNGLSIKWTASN-----EQIOVGADG--VKFAMVNGVVGAGIDGTTTRTRDEIGF 1083  
 Db 961 VKSGETVAVPVLTVSNVPGIATAELOVGFEDATLLEVASITVGD-----VLNPVNF 1011  
 QY 1084 TG-TNGSLDK-----SKPHLSKDG-----INAGGKKTINIOGGEIAKNSHDAVTGCKI 1130  
 Db 1012 SSVNGSTIKLLFLDDTLGSLISKDGVLATINEKAKTVTSKVTTPVAVSGTVPFADGTL 1071  
 QY 1131 YDLKTELENKISSTAKTAQNSLHIESVADEOG-----NNFTVSNPSSYSDTSKTSDVITF 1185  
 Db 1072 ----ABELNKTVAGSVTIEPSPQVKVTVATVGTATVKSGETVAVP-----VTL 1115  
 QY 1186 AGENGITTKVNGVVRVGDIDTKGLTTPKLVGNNGRGIVIN-SQNGQNTITGLSNFLA 1244  
 Db 1116 SNVPGIAT-----AELOVGFEDATL-LEVASITVGD-----VLNPVNFSSVNGSTIKLL 1165  
 QY 1245 NVTNDRGVRRTEQGNIIKDEKTRASIVDVLNAGFNLOGNGEAVDFVSYDVTNFPANG 1304  
 Db 1166 FLDDTLGSLISKDGV-----VLA-----TINP-KA 1189  
 QY 1305 NITTAQVT-----YDTSKTSKVYDVNVDDTTIEYKDKLGVKTTTLTSTG-- 1351  
 Db 1190 KTVTSKVTTPVAVSGTVPFADGT--LAELKYETVAGSVTIEPSPQVKVTVATVGTATGKV 1247  
 QY 1352 --TGANKFALSN-----QATGDA-LVKASDIIVAHNLTLGSDIOTAKGASQANNAG 1399  
 Db 1248 GETVAVPVLTVSNVPGIATAEVQVGFEDATLLEVASITA-----GDIV---LNPSVNFSS 1297  
 QY 1400 YVDADGNKVTY-DSTONKYYOAKNDG---TVD-KTRKVAADKL--VAQAQTP---DGTILA 1449  
 Db 1298 VVNGSTIKILFLDDTLGSLISK-DGVFAITNFKIAPSTGTTPVAISGTPVFAFGTLA 1356  
 QY 1450 QMNYKSVIN-----KEOVNDANKQG-----INEDNAFVKGLEKAASDN 1488  
 Db 1357 EVQYKTVAGSVTIAADIKAVKATVGTATGKAGDTVAVPVLTVSNVSGIATVELQLSFDAT 1416  
 QY 1489 KTRKNAVTGDLNVAQTPLTFAG-DTGTTAK-----KLGETITIKGGQTD-----NKL 1537  
 Db 1417 LLEVASITAGDI--VLNPVNFSSVNGSTIKILFLDDTLGSLISKDGVFATVNFVKAS 1474  
 QY 1538 TDNNITGV-----VAGT---DGTVKLAKD-----LTNLSVNVNAGGTYKIDEGISFVDANG 1584  
 Db 1475 TATNSAVTPTVSGTVPFADGTLAEKLSESAAGRLTILPTVIVVDSTVAPVATVTEFKANQ 1534  
 QY 1585 QAKANTPVLNGLDLGKGVSNVKGKTDDAANVOQLNEVRNLLGLGNDNADGNQVNI 1644  
 Db 1535 ADAALITMTLNGTTFSAIKNGTATLVKGTDTYIVSENVTISK----- 1575  
 QY 1645 ADIKKDPNPGSSSNRTVIAKGTV-----LGGKGNNDTEKLTATGGVQGVYDKDGNANGDLSN 1700  
 Db 1576 AVLAK-----QTGVTTLFEEFDRGNSAKVVAVVAYKEIQI-----VNSTITP 1615  
 QY 1701 VVVTKQDGSKKA-LLAYNNAAGOFNYVTNPAEALDRINEQGIREFHVDNGOEPVYQG 1759  
 Db 1616 VVATFEKTAAKQADVVTVMNLNGNT-----FSAIKNGTTLV--- 1652

Db 63 GLSHNOYQDNVNPQGNVLNNSREAGLSQLAGQIGANPNLNGREASVILNEVIGRNPSSL 122  
 QY 304 -----ADQYKATROGSDTDIFSIGNNNNNNSIRKIIINVGA-----GSRDTD 346  
 Db 123 HGQOEIFGMAADYVLANPNCISQSCGFINTSHSLVWGNPLVENGVLQGYSTFGNRTL 182  
 QY 347 AVN-VAQLKLEELANRKTIFKGD-----GDNNSNSVERGLG-NFLTKIGDAQTN 394  
 Db 183 SLNGTLNAGGVLDLAPKIDSRGEVIVQDPKQNGKVTSAAINAISGLNKRVARDDGTVAAS 242  
 QY 395 -----ALTEANIGVYTDGNGKLKAKELTG-----LTSVSATNKIT 431  
 Db 243 QOMPTALDSYLLSQMAGRINIINTAGSGVKLAGSLNAGDELKRAYDIRSERVDDAS 302  
 QY 432 VSNNTNNNAELQGGITFSPITGKTDKTVYISIDGLKFTNDSNSIATGTRITKKKIGF 491  
 Db 303 SNKNGGDYQNYRGGI-----YVNDRSSQTLTRTELKGNISL 341  
 QY 492 AGTNDGVDESKPYLDNEKLVGNSTLNSGSLTVNN-----TTG 529  
 Db 342 VADNHA-----HLTATDINGEDITLOGGKLTLDGQOLKOTQGTDDRWFYSWQYDVIRE 395  
 QY 530 NKQIOLVGANGIKFATVANNVANTSATGTARI--TEEKIGFACTNDGVDEQ----- 578  
 Db 396 REQLQ-----QAGSTVAASGSAKLISTQEDVKLLGANVSADRALSVAARDV 442  
 QY 579 --APYLDKERLKVREITTDGINSAGNHKITGLTINGIANTD---AVTIKQLK-DAKPTL 632  
 Db 443 HLAGLVEKDK-----SSERGYQR-NHTSSLTGRWNSDESELSKASELSEGETLT 493  
 QY 633 NAGDGSINSNGDLVDSSGNIITPYNISVKTILNSGTSNKNKFSVSNADNNSLVT 692  
 Db 494 KAGRNV---STQAKVHAQRDLT-----IDADNQIVGVQVQRTANAKAV 533  
 QY 693 AKDLADYLNKVNETADALSALPFSKVQNGDNNNA-----ITVGKD-TNGKPTENTLKLKENG 747  
 Db 534 RDKTSWGG-----IGGDNKNNSNRREISHASELTSG---GTLRLNGQQG 576  
 QY 748 VNITTRATGTVTFGIDQSNGLTTPKLTIVGSDTNGRLVIEQVPSADGNSTKNIILGLSP 807  
 Db 577 VTITGSKARG-----QKGEVTA-----THGG-----LRIDNALST 607  
 QY 808 TLPASIPSPERNALGNTIEEKDKSNAASIDDLVNLNAGFNLK-NNGKDKDFV-STYDTVDF 865  
 Db 608 TVDKIDARTATPNTSSSHKADNSQSSPASELSKSDTNLTLSHKDADVIGSQVASGGE 667  
 QY 866 IDGNATTATVYDEANQTSKAVDVANVDEKTIETLDGNGKKQLGVKTIK--LLETSTNGN 923  
 Db 668 LSVESKTNINVKABERQ-----NIDEQKALTVNGYAKEAGDKQYRAGLRIEHTRDS 721  
 QY 924 ATTFSTDDHVALYKASDIAGNLNTLAEIHTTKGTANTALQTFVTKVDENDKADDTNAI 983  
 Db 722 EKTTRTEN-----SASSLSG----- 736  
 QY 984 TVGKDGTSKVNLTCLKGK-----NGLDIKDKGTVTFGINTQSGLKAGDSTTLNNGL 1038  
 Db 737 -----GSKVLKAEKDVTFSGSKLVADK-GDASVSGNKVSVSLAADDKTASN--- 780  
 QY 1039 SIKNTASNEOIQVAGDGVKPFAMVNGVVGAGIDGTTTRITRDEIGFTGTNGSLDKSKPHLS 1098  
 Db 781 -----TEQTKGG-----GFFYT-GGIDK-----L 799  
 QY 1099 KDGINAGGKKTNIQSGEIAKNSHDAVTGKIIYDLKTELENKISSTAK-TAQNLSLHEFSV 1157  
 Db 800 GSGVEAG-----YENNKTOAQSSKAITSGS--DVAGNL--TINARDKLTQOQAQH--SV 847  
 QY 1158 ADQGNFTVSNYSYDTSKTSDDVITFAGENGITTKVNGKV-VRVGIDOTKGLTTPKLT 1216  
 Db 848 GGAYOENAGVVDHLAADTASTT-----TTKTDGVNIGANVDYS-AVTRP--- 892  
 QY 1217 VGNNGKGIIVINSONGONTITGLSNTLANVTNDKGVSRITTEQGNITTKDEKTRAASITDV 1276  
 Db 893 VERAVGAAKLDTATGVINDIGGIG--APNVGLDIGA-----QGGSEKSSSSQAVVSSV 945

QY 1277 LSAFENLQNGEAVDFVSTYD---TVNFA--NGNTTTAKVTYDDTSKTSK-----VYYD 1325  
 Db 946 QAGSIDINAKGEVRDOGTQYQASKGAVNLTDASHRSEAAANRODEQSRDTRGSAGVRVT 1005  
 QY 1326 VNVDDTTIEVKDKLGVKTTTLTSTGTGANKFALSNGATGDALVKASDIVAHLNLTSGDI 1385  
 Db 1006 TTGSDTLTVDAK-----GEGGTQ--RSNSSASQA-----VTGSI 1036  
 QY 1386 QTAGASQANNSAGYVDADGNK-VIYDST---DNKYFQAKNDGTVDKTKFVAKDKLVAQA 1441  
 Db 1037 DAANG-----INVVKDAIYOGTALNGRGTAVNAGG-DIRLQOASDK---QS 1082  
 QY 1442 QTPDGLTLOMNVKSVINKEQVNDANKKOGINEDNA-FVKGLEKAASDNKTKNAAVTVGDL 1500  
 Db 1083 ESRSG---ENVK-----ASAKGFTADSKNFGAGFGGGTHNGESSSTAQVGN 1128  
 QY 1501 NAVAQPLTLAGDTGTAKKLGETLTIG---GQDTNKLTDNNIGVAGTGDFTVKLA 1556  
 Db 1129 -----SGOQVEL-RAGRDLTLQGTVDKVSQGDVLSAGNKVALQAAESTQTRKES 1177  
 QY 1557 KDLTNLNSVNAGGKIDEGISFVDANGQ---AKANTPVLISANGLDLG--GKV-ISNVGK 1610  
 Db 1178 KLSGND-LGASGDSKEKTGGNLSAGGAFDIAKNESATERQGATTASDGKVTLSANGK 1236  
 QY 1611 GTKDTPDAANYQ-----QLNEVRN-----LLGLGNDNADNGNOYNIA 1645  
 Db 1237 G---DDALHLQAGKAVSGGSAALEAKNGGILLESKAKNEQHKDNWSLGI-KANAKGGQTFNK 1292  
 QY 1646 DI--KRDPSNGSSSNRTVIKAGTVLGKGNNDTEKLTG-----GVQV 1686  
 Db 1293 DAGGKVDNPFGKDTHT--TLGAGLVGVEQDDKTHANTGITAGDVTLSNGKDTRLAGARV 1350  
 QY 1687 GVDK-DGNANGDLSNVVVKTKD---GSKKALLATYNAAGOTVYTNPNPAEI--DRINEQ 1741  
 Db 1351 DADSVQGVKGGDLH---VESRKDVENGCVKDV-----DAGLSH--SNDPSSSITSKLSKV 1400  
 QY 1742 GIRFFHVNDGN-DEPVVQGRNGIDSSASGKH-SVAIGFOAKADGAAVAIGR----- 1791  
 Db 1401 GTPRY---ACKVREKLEAGVKNVADATTDKYNVARRLDPQDITGAVSF-SKABGKVTL 1457  
 QY 1792 QTAGNOSIA-IGDNAQTGDSIATGIGNVAVAGKHSAGIDPSTVKADNSYSVGNKNQF 1850  
 Db 1458 ATTAGKPPQGLMDRGART-----VGGAVKDSITGPAGRQGLHKV--NADVYNNN-- 1505  
 QY 1851 TDATQTDVFGVGNITVTESNVAL--GSNSAISAGTHAGTQAKKSDG----- 1896  
 Db 1506 -----AVGESATAGKNGVALQVGGQTLTGGEIRSQQKVELGGSOVSQDYNQ 1556  
 QY 1897 --TAGTTTTAGAT--GTVKGFAGQTAGVAVSVGASGAEIRRIQNVAAGEVSATSDAYNG 1951  
 Db 1557 RYQGGGRVDAATAVGGLLGGAAKQSVAGNVP-----FASGHASTQQADAKAG 1603

RESULT 14  
 FHAB\_BORPE  
 ID FHAB\_BORPE STANDARD; PRT; 3591 AA.  
 AC p12255;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE FILAMENTOUS HEMAGGLUTININ.  
 GN FHAB.  
 OS Bordetella pertussis.  
 OC Bacteria: Proteobacteria; beta subdivision; Alcaligenaceae;  
 OC Bordetella  
 OX NCBI\_TaxID:520;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE-90355839; PubMed-2388559;  
 RA Rellman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;  
 RT "Genetic characterization of Bordetella pertussis filamentous  
 haemagglutinin: a protein processed from an unusually large

precursor. ;  
RT Mot. Microbiol. 4:787-800(1990).  
RN SEQUENCE OF 1-3261 FROM N.A.  
RP MEDLINE=9202384; PubMed=2539596;  
RX Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;  
RA "Filamentous hemagglutinin of Bordetella pertussis: nucleotide  
RT sequence and crucial role in adherence";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).  
CC -!- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND  
CC INFECTION.  
CC -!- SUBCELLULAR LOCATION: SURFACE.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M60351; AAA22974.1; .  
DR EMBL: M60351; AAA22975.1; ALT\_INIT.  
DR EMBL: M60351; AAA22976.1; ALT\_INIT.  
KW Antigen; Hemagglutinin.  
SQ SEQUENCE 3591 AA; 367420 MW; EF7418B3D6E5138 CRC64;  
  
Query Match 3.9%; Score 404; DB 1; Length 3591;  
Best Local Similarity 20.1%; Pred. No. 3.2e-07;  
Matches 487; Conservative 282; Mismatches 904; Indels 750; Gaps 110;  
  
QY 9 FNKATGTFMAVEAYKASHSTGG---GSCATGQVGSVRTLSFARIAALAVLVIGATLNGS 64  
DB 264 YDHATRRATPINAGAAAGAAVAIDGTAGAMYGKHIL-VSSDSGLGVRQLGSLSFPS 322  
QY 65 AYAOQITTKIEIGOTKINNTLKGDALATCEASIAFGSLSKAQSQAIAIGSVKPDNNG 124  
DB 323 AITVSSQGEALGDA---TVQRPLSLKAGVVSAGKLASGGGVNAGGVAKIASAS 378  
QY 125 SNGVNGSHAKNESIAI---GGDVLAEGDASIAIGSDLLYLPKNLDLKNFEKHLIHGHEI 181  
DB 379 SVGNLAVOGGKQVATLLNAGGTLVSGRAVOLGA-----ASS 417  
QY 182 LKKITQTSDDGRIK-----YRRTRAQGHASTAVGAMSYAAGHFSNAPGYATAEAYSLA 235  
DB 418 QALSVNAGGALKADKLSTRVRDVGKQAVLGSAS-----SNAL-----SVR 461  
QY 236 VGLAAQATKQS-----STAVGSNAKANAFATAIG---GNTVNLGRGVALGFG 281  
DB 462 AGGALKAGKLSATGRDLDVGRQAVTLGVSASDGLSVSAGGNLRANELVSSAQLEVRGQR 521  
QY 282 SQILDRDNTDASAVPLGKTLADQYKATRGDSTDIFSGNSNNNNSSIRRKLIINVA- 340  
DB 522 EVALDASSARGMTVVAAGALAA-----RNLSQKAI 553  
QY 341 GSRDITDAVNAQLKVLVEELARKKTFKGGDNNNSNVERGLNTLTIKGDAQTNALTEAN 400  
DB 554 GYOGGEAVSVANANSDAEL---RVRRGQVDLHDLASAARG---ADISGEGR-----VN 600  
QY 401 IGVVTDGKGLVKKL---AKELTGLT-----SVSATN-----KITVSNTNN--NNAE 441  
DB 601 IGRARSDSVKVAHGALSIDSMTALGAIGVQAGGSVSASAKDMRSRGAVTVSSGGAVNLGD 660  
QY 442 LOSGGITFSPIGTGKT-----DKTVYSDGLK--FTDNSNSTATKGTTRITKKIGF 491  
DB 661 VQSDGQVATSGAMTVRDVAAAADLALQDALQAGFLKSAGAMTVNG-----RDAVRL 715  
QY 492 AGTNDGVDESKPYLDNEKLK---GNSTLNS-----GSLTVNNTTGNKOIQVGANGIKPAT 544  
DB 716 DGAHAG-----GQLRVSSDGOALGSLAARGELTV-----SAARAAT 752  
QY 545 VA---NNVANTSATVGPARTITEKIGFAGTNGVDQEQAYLDKERLKVGRVEITTDGGINA 602

Db 753 VAEKSLDNIISVTGG-----ERVSQSVNSA---SRVAISAHGALDVGKV--SAKSIGL 802  
QY 603 GNHKITGLTN-----GIANTDAVTIKQLKD-AKPTLNAGDGISINSNNGDLVDSSGNI 654  
Db 803 ECWGAVGADSLGSGDAISVSRDQAVRVDQARSADISLARGGATLGA-----VEAAGSI 857  
QY 655 TPTPTNISVKTTKLNSNGTSGNNKFSVSNHNNLSL-VTAKDLADYLKVNKNETADSLPS 713  
Db 858 -----DVRGSGTVAANSLHANRDVRVSGKD-----AVRVTATSGGG 894  
QY 714 FKVQNGDNNNAITVCKDTNGKTFNTKLKNGGVNITNRTATGTFGIDQSNGLTTPK 773  
Db 895 LHVSSGRQLDLGAQAR-----GALALDGGAGVALOSAKASGTLHVQCGEHLDLGTLA 947  
QY 774 LTVGSDTNGN-----RLVIEQVPSADGNSTKNIIGLSPT-----LPSTAS 814  
Db 948 AVGADVNGTGDVRVAKLYSD--AGADLQAGRSMTLGIIVDTTGDILQARAQOKLELGSYKS 1005  
QY 815 PSGRNIALGNTE-----EKDKSNAASIDDDVLNAGFNLKNGK--- 852  
Db 1006 DGLQAAAGGALSAAAEVAGALELSGGQVTVDRASARRIDSTGSGVIGALKAGAVEA 1065  
QY 853 -----DKDFVSTYDTVDPIGDNATTATVYDEANQTSKVAIDVNVDEK---TIELT 900  
Db 1066 ASPRRARRALRQDEFTPGSVVRAQGN---VTVGRGDPHQGVLAOGDIIIMDAKGGTLLR 1122  
QY 901 GDNGKKQLGVKTKLTETSTNGNATTFSTDD---DIALVKASDIAGNLNTLAEETHTKG 957  
Db 1123 ND-----ALTENGT-----VTISADSAVLEHSTHESK-----ISQSVLAAG 1159  
QY 958 TANTALQTFVKKVDENDKADDTNATVVGKDGTSKVNLTKLKNGKGLDIKTDK-----DG 1013  
Db 1160 D-----KGPASVSVKAKKFLNG 1178  
QY 1014 TVTFEGINTQSLKAGDSTTLNNCLSLKNTASNEQIQVGADGVKFAWNVNGVVGACIDGT 1073  
Db 1179 TL-----RAVNDN---NETMSQRQIDV-----VDGR 1201  
QY 1074 TRITRDEIGFTGTGSLDKSKPHLSKDGINAGGKIKTNIQSGETAKNNSHDAVTVGGKIYDL 1133  
Db 1202 PQIT-----DAVTGEARKDESVDAAALVADGGPIV-VEAGELVSHAGGIGNG----- 1248  
QY 1134 KTELENKISSTAKTAQNSLHE-FSVADEQG-----NNFTV-SNPYSYDTSKTSDV 1182  
Db 1249 -RNKENGASVTVRTTGNLVNKGYSAGKQGVLEVGGALTNFELVSGDGTORIEAQRIENR 1307  
QY 1183 ITF-----AGENG-ITTKVNKGVVVRGIDQTKGLTTPKLTVGNNGKGIIVNSONGONTI 1236  
Db 1308 GTFQSQAPAGTAGALVVKAAEAIVHDGVMAKKG---EMQIAGKGGSPVTVAGAKATTS 1364  
QY 1237 TGLSNTLANVTN-----DKGSVRTT-----EOGNIKDEKTRASIVDVLNAGFNLQ 1284  
Db 1365 NKLSVDVASWDNAGSLDIKKGGAQVTVAGRYAEHCEVSIQGDYTVSA---DAIALAAQVT 1421  
QY 1285 GNGEAVDFVSTYDTVNFAN-----GNITAKVTYDDTSKTSKVYVYVNV 1328  
Db 1422 ORGGAANTLSRHDIT-RFSNKLRLMGPLOVNGAGGPVSNTGNLKVREGVT--VTAASF--- 1475  
QY 1329 DDTTTEVKKDLGKVKTTTLTSTGTGAN--KFALSNOATGCDALVVKASDIVAHLNLTSGDIQ 1386  
Db 1476 NETGAEM-----AKSATLTTSCAARNAGKMOVKEAAT---IVAAS--VSNPGTFTA--- 1522  
QY 1387 TAKGASQANNAGYVDADG---NKVIYDSTD---NKVYQAKNDGTVDKTEV----- 1432  
Db 1523 ---GKDIPTVTSRGGFFNEGKMESNKDVIKTQFNSGRVLDKADHDLTVTASQADNRGSL 1579  
QY 1433 -AKDKLVAQAQTPD--GTL-----AQMNVKSVINKEQVNDANKKOGINEDNAFVKGLEKA 1484  
Db 1580 KAGHDTVQAQRIDNSGTMAAGHDATLKAPHLRNTGOVVAGHDIIHINSAKLENTGRVDA 1639  
QY 1485 ASDNKTNAAVT-VGDLNVAQAQTPTLTFAGDT-----GT---TAKKLGFT 1524



Db 1640 RNDIALDVADFTNTGSLYAEHDAITLTAQGTORDLVDDHILPVAEGTLRVKAKSL--T 1697

Qy 1525 LTIKGGQ-----TDNKLFDNNIGVAGDGTGTVKLAKDLT-----NL-NSVNA----- 1567

Db 1698 TEIETGNPGSLIAEVOENIDNKQAIIVG-----KDLTSSAHGNVANEANALLWAA 1748

Qy 1568 -----CGTKIDKGIISFVDNCAQKANTPVLSANGL-----DLGKVISNVGKGTG 1613

Db 1749 GELTVKAQNTITNKRAALIEAGGNARLTAVALLNKLGIRAGEDMHLDPRIENTAKLSG 1808

Qy 1614 DTDAAANVQOL--NEVRNLGLGNDN---ADGNQVNIADIKKDPNSGS--SSNRTVIKAGTV 1667

Db 1809 EVQRKGVQDVGGEGHCRWSSGIVYVWLRAGNGKAGTAAAPWYGDLTAEGSLIEVGKD 1868

Qy 1668 L-----GGKNNDEKATGGVGVGVDKGNANGDLSN----- 1700

Db 1869 LVINAGARKDEHRHLLNEGVIQAG--GHGIGGDVDRNSWRTVSAMEYFKTPLPVSLTA 1926

Qy 1701 -----VW-----VKTQD----- 1708

Db 1927 LONRAGLSPATWNFOSTYELLDYLLDQNRKYEYIWLGYPTTEWSVNTLKNLDLGYQAKPA 1986

Qy 1709 -----GSKKALLATYNAAGQTNVYNNPAEADRINE 1740

Db 1987 PTAPPMPKAPELDLRCHTLESAGRKIFGEYKKLQGEYKAKMAVQAVAYGEATRRVHD 2046

Qy 1741 Q-GIRFFHYNDGNOEPVVGGRNGIDSSASGKHSVAI--GFQAKADGEA----- 1785

Db 2047 QLGQRYGKALGMDAETKEVDGLIQEFAADLRTVYVAKQADQATIDAETDKVAQRYKSQID 2106

Qy 1786 -----AVALGROTAQGNQSIAGIDNAQATGDSIAIGTGNVAGKHSUGAIGDPSTVKADN 1840

Db 2107 AVRLQAIQPGRYTLAKALSAALGADMRALGHLSQLMQWRKDFKAGKRGAGIA----- 2157

Qy 1841 SYSVGNNOFTDATQDVFVGNNITVTESNVALGSNSAISAGTHAGTQ--AKKSDGTA 1898

Db 2158 -----FYPEQET-VLAAGAGLTLS-NGAIHNGENAAQNGRPEGKIGAHSAITSVS 2206

Qy 1899 GTTTAGATGTGKGFAGQAVGAVSVG-----ASGAERRIONVAAGEV-----SATST 1946

Db 2207 GSFDAIRDVGLKRLDIDDAALAVLNPHIFTRIGAQTSLADGAAGPALAQARQAPET 2266

Qy 1947 DAV-----NGSOLYKATQSIANATNLDHRHQNENKANAGIS--SAMAMASWPQAYIQC 1999

Db 2267 DGMVDARGLGSADALASLASLADAAQGLEVSRRNAQVADAGLAGPSVAAPAVGAADV-G 2325

Qy 2000 RSMVTGGIATHNGQGVAVVGLSK 2022

Db 2326 VERVTG---DQVDQPVVAVGLEQ 2345

RESULT 15

P2P\_LACPA

ID P2P\_LACPA STANDARD; PRT; 1902 AA.

AC Q02470:

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-

DE ASSOCIATED SERINE PROTEINASE) (LP151).

GN PRTP.

OS Lactobacillus paracasei.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

OC Lactobacillus.

OX NCBI\_TaxID=1597;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCDO 151;

RX MEDLINE=92381481; Pubmed=1512565;

RA Holck A., Naes H.;

RT "Cloning, sequencing and expression of the gene encoding the cell-

RT envelope-associated proteinase from Lactobacillus paracasei subsp.

RT paracasei NCDO 151."

J. Gen. Microbiol. 138:1353-1364(1992).

[2]

RL SEQUENCE OF 189-196.

RN MEDLINE=92226694; Pubmed=1564442;

RX Naes H., Nissen-Meyer J.;

RA "Purification and N-terminal amino acid sequence determination of the

RT cell-wall-bound proteinase from Lactobacillus paracasei subsp.

RT paracasei."

RL J. Gen. Microbiol. 138:313-318(1992).

CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE

CC GROWTH OF THE BACTERIA ON MILK.

CC -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD

CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,

CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND

CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS, AND

CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED

CC INSULIN B-CHAIN.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE

CC SUBTILASE FAMILY.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M83946; AAA25248.1; ..

DR PIR; B44858; B44858.

DR HSP; Q99405; IMPT.

DR MEROPS; S08.019; ..

DR InterPro; IPR002029; ..

DR InterPro; IPR001899; ..

DR Pfam; PF00746; Gram\_pos\_anchor; 1.

DR Pfam; PF00082; Peptidase\_S8; 3.

DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS00136; SUBTILASE\_ASP; 1.

DR PROSITE; PS00137; SUBTILASE\_HIS; 1.

DR PROSITE; PS00138; SUBTILASE\_SER; 1.

DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; FALSE\_NEG.

KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal;

KW Transmembrane.

FT SIGNAL 1 33 POTENTIAL.

FT PROPEP 34 187 POTENTIAL.

FT CHAIN 188 1902 PII-TYPE PROTEINASE.

FT DOMAIN 188 1876 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1877 1895 MEMBRANE ANCHOR (BY SIMILARITY).

FT DOMAIN 1896 1902 CYTOPLASMIC (POTENTIAL).

FT ACT\_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE

FT PROTEINS.

FT SEQUENCE 1902 AA; 200253 MW; D8C9F38CEE5DA582 CRC64;

Query Match 3.9%; Score 399; DB 1; Length 1902;

Best Local Similarity 20.7%; Pred. No. 2.3e-07;

Matches 450; Conservative 249; Mismatches 880; Indels 592; Gaps 100;

Qy 31 GSCATQGVGSVRTLSF-----ARIALAVLVIGAT-----LNGSAYAQITTK-I 74

Db 19 GALAVLPVGEIQAKAISQOTKYVSSSLANTVKAATAQKATDTTAAATNQIAATLAAKGI 78

Qy 75 EIGQTNKI-----NNTLKGDALATGEASIAFGSLKSAQGSIAAI 114

Db 79 DYNNKLNKVOQDQTYVDVIVQMSAAPASENGTLRTDYSSTAEIQOETNKVIAAQAQSKAAV 138

Qy 115 GSVKPPDPNNGSNV-----GSHAKGNESIAI- 141

Db 139 EQVTOQTAGESYGVVNGFSTKVRVVDIPKLQIAGVKVTVTLAKVYPTDAKANSMANVQ 198

QY 142 -----GGDVLAEGBASIAIGSDLYLPLKNDLKNFHLIHGHEILKIKIOTSDG 191  
Db 199 AWMNXYKKGEGTVVSDVIDGIDTHKMDRLSDDKV-----LKYDVEKEDTDIAKHG 252  
QY 192 -----KIKYRRTAAGHASTAVGAMSYAQC--HFSNAFGTYATABAAYSYLAAGLAAQA--- 242  
Db 253 RYFTYSKVPYGFYANDNDTITDDTVDSQHGHHVAGIIGANGCTGDDPTKSVVGVGAPEAQLL 312  
QY 243 -----TKQSIAGVSNKAKANAFAATAIGNTVVNLGRGVALGFGSAILDRDNNNTDASAY 296  
Db 313 AMKVFTNSDTSATTGSAFLVSAIEDSAKIGADVLNMSLG--SDSGNQTL----- 360  
QY 297 VPLGKTADQYKATROGDSTDIIESIGNNNNNNSIRKKLIN-----VAGASRDT 345  
Db 361 ---DPEIAAVONANESGTAA--VISAGNSGTSATQ--VNKYDYGLODNEMVGTPTCSR 414  
QY 346 DAVNVAQLKVEELANRKITFKGDCDNNNSVERGLGNTLPIKGDATONALTEANIGVVT 405  
Db 415 GATTVASAENT-DVISOAVTIT-DGKDLOLQGP-----TIQLSSNDEFTSGFDQKKFYVK 467  
QY 406 DGNGLKVKLAKELGTLTSVSATNKITYSNTNNNNELOSGLTFSPITGKTDKTVYS-- 463  
Db 468 DASG---DLSKGAADYTTADAKGTAI-----VARGELNFA-----DKQKTAQA 508  
QY 464 --IDGLKFTNDSNSIATGTTTIT-----KKKIG-----FAGTND--GVDESKPY 504  
Db 509 AGAAGLIIVNDGTATPLTSIRLTTFEGLSSKGTQKGLVDWVTAHPDDSLGVKIALTL 568  
QY 505 LONEKLVGNSTLNGSLTVVNTTGNKQIOVGANGIKPATVANNVANISAT-----VG 557  
Db 569 LPNQRYTEKMSDFTSYGPVSNLSFKPDITAPGGNINWSTNNNGTSMASPPFTAG 628  
QY 558 TARITEEKIGFAGTNDGVEQAP-YLDERLK-----VGRVEITDSDGINAGNHKITG 609  
Db 629 SOALLUQALN-----NKNPFTADYKQLGTAITLDFLKTVMNTAQPINDINY----- 676  
QY 610 LUNGIANTDVAITIKOLDAKPTLNAAGDISINNSNDGLVDSG--NITTPYINISVKTWK 667  
Db 677 --NNVIVSPRGAGLVQVKAIDA-----LEKNPSTVVAENGYPAAVELKDFSTDTKTFK 729  
QY 668 LN-SNGTSONKFSVNAHNNLSITAKDLAD---YLNKYNETAADSALPSFKVONGNS 722  
Db 730 LFTNKTTHLTYQMSNTDITNAVYTSATDPSGVLYDKKIDGAAIKAGSDITVPAGKTA 789  
QY 723 NNAITVG--KDTNGKTF--NTLKLKENG--VNITNTRATGTVTFG--IDQSNGL----- 769  
Db 790 QIEFTLSLPSKFDQOQFVBEFLNEFKGSDGSRNLNLYPMGFGDNDGKIVDSLNGITYSPA 849  
QY 770 -----TTPKLTVGSDTNGNRLVIEQVPSADGNTKNIILKGLSPTLPSIASPSGRNIALGN 824  
Db 850 GGNYGTVPLLT--NKNTHGQYGGVMYTDADGKOTVDD-----QALAFSSDKN----- 894  
QY 825 TIEBKDSNAASIDDDVLNAGFLKNNKDKDFVSTYDITVDFIDGNATTATVYDEANQTS 884  
Db 895 -----ALYNDLSMOYLLRN-----ISNVQ--VDILDGOGNKVTTLSSTNQOT-- 935  
QY 885 KVAYDVNVEKTIELTGDNKKKOLGVKTKLTETSTNGNATFTSTDDHIALVKASDIAGN 944  
Db 936 KTYIDAH-----SOKYIYINAPAMDG---TYIDORD-----GN 965  
QY 945 LNTLAEIHTTKTANTALQTFVKKVDE-NOKAD-----DTNATVKGDTSGKVN 995  
Db 966 IK-----TADGGSYTRISVPEGGDKRQVDFVPFKLDSKAPTVRHVALSAKTE 1014  
QY 996 TLKLGKNGLDIKTDKG-----TVTFGINTOSGLKA--GDSTTLNNGLSIKNTASNQ 1048  
Db 1015 NGKTYLLTAEAARDLGLSLDATKSVKATNEVNTLNDATETDAGTADGYTKIETPLSDRQ 1074  
QY 1049 IQVGADGVKFAWYNNGVVNGAGIDGTTIRIDEIGFTGNTGSLDKSKPHLSKDG--INAGG 1106  
Db 1075 AQA-----LGNG-----DNSAELYLTDNASNATNODASVQPKPGSTFSLIVNGGG 1119

QY 1107 --KKITNIOGSEIAKNSHDAVT--GGKIYDOLKTELENKISSTAKTAONSILHPSVA-DEOG 1162  
Db 1120 IPBKISSITTTG-----YEANTQGGCTYTFSGTYPAADVDTYDAQGGKHLNDLNTTYDAAT 1173  
QY 1163 NNETVSNPSSYSDTSKTSDVITFAGENG-----TTKVNKGVVVRVGIDQTKGLTTPKLT-V 1217  
Db 1174 NSFTASMAVNTADYAAQVDLYADKAHTQLLKHFDTKVR-----LTAPITFDL 1220  
QY 1218 GNNNGKGIIVNSONGQNTITGLSNTLANVNDKSGSVRTTEQGNIIKDEKTRAASTVDVL 1277  
Db 1221 KFNNG-----SDQTSEATI-----KVTGTVSSD--TKTVNVGDTVAALDAQHHFSVDVPV 1268  
QY 1278 SAGFNL-----QGN--GEAVDFVSTYD-----TVNFANG-----NITTTAKVY 1313  
Db 1269 NYGDNITKVTATDEGNTTTEQKTIITSSYDDPVLKNAVTFDQGVKFGANEFNATSARF-Y 1327  
QY 1314 DDTSKTSKVYDVNVDTTIEVKDKKLGVR-----TTTTSTGTGANKFAL-----SNQ 1362  
Db 1328 DPKTGATITCKVKHPTTTLQVQKQISIKNDLTFSTLDLGLTGOKPGVGVVGDITQNK 1387  
QY 1363 ATGDALVKASDIVAHLNLTLSGDIOTAKGASQANNSAGYVDADGNKVYYSDTKNYQAKN 1422  
Db 1388 TFOEALTFILDAVA--PTLSLDSST-----DAPVYTNDFNFQITGTATDNAOYLS-- 1435  
QY 1423 DGIWDTKTEYAKDKLVAAQAOTPDSTLAQNMVKSVINKEQVNDANKK--QG1NEDNAPVK 1480  
Db 1436 -----LAINGSIVASQYADININSGKPGHMAIDOPVKLLEKN-----VLT 1476  
QY 1481 LEKAASDNKTNAATVYVDGLNVAQAOTPLTFAGDTGTTAKKLGTELTIKGGOTDNKLTDN 1540  
Db 1477 VAVTDSENNITTKKITV-----YVEPKKTLAAPIVTPS-----TTEPAKTVT--LTAN 1522  
QY 1541 NIGVVAAGTGTGVLAKDLNLSVYVAGG--TKIDEKGLSFVDANGAKANTPVLSANG 1597  
Db 1523 -----AAATGETVQVSAD-----GGKTYQDVPAAGVT-VTANGTFFKSTDDLYGNE 1567  
QY 1598 LDLGKGLSVSNVKGKTDKTDAAVQOQLNEVRNLLGLGNDNADGNQVNIADIKKDPNSGSS 1657  
Db 1568 SPADVYVVTNI-----KADDPQAQLOTAQALTNL-----IASAKTLSASGKYD 1610  
QY 1658 NRTVIKAGTVLGGKGNNDTEKLTAG--GVQGVQDKGNANGDLNWNVVKQKDGSKKALL 1715  
Db 1611 DATT-----TALAAA-----TOKAQTALDQTOASVDSLTCANHDLQT-----AINOLA 1653  
QY 1716 ATYNAAGOTNVYTNPAEAIIDRINEQGLRPFHVNDGHOEPVWQGRNGIDSSASGKHSVAI 1775  
Db 1654 AKLPADKKTSL-----NQLQSVKKAALGTDGLGNQDTPSTGK 1689  
QY 1776 GFQAKADGEAAVAJGROTQACNQSIATIGDQAATGDSOSIAIGTGNVAGKHSGAIGDPST 1835  
Db 1690 TFTAALDDLVA-----QAQAGTQ-----ADQLQA-----SLAKVLDVAFLAKLAEG----- 1730  
QY 1836 VKADNSYSGVNNQOTDATQTDVEGVNNITVTESNSVALGNS-----SAISA 1883  
Db 1731 IKAATPAEYVGNK-----DAATGKTWTADIAOTLTSGQASADSKLAHLQALQSLKTKVAA 1787  
QY 1884 GTHAGTQAKKSDGTAGTTTTAGATGTVKGFAGTAVGAVSYGASGAEERRIONVAAGEV-- 1941  
Db 1788 AVEAAKTAGKDDITGTSKGGGGTTPAPAGDT-----GKDKGDEGSQPSGGNIPT 1840  
QY 1942 -----SATSTD 1947  
Db 1841 KPATTITSTSD 1851

Search completed: September 13, 2001, 12:53:37  
Job time: 847 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:55:10 ; Search time 112.99 Seconds  
(without alignments)  
2403.949 Million cell updates/sec

**Title:** US-09-361-619-9  
**Perfect score:** 10356  
**Sequence:** 1 MNHIIKVFINKATGTGTFMAVA.....NGSADTQGHVCAAVGAGGFHF 2053

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 s

```
Database :
SPTRMBL_16.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_unclassified.*
13: sp_vertebrate.*
14: sp_virus.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	1206	11.6	2353	2	P71401	haemophilus
2	1191.5	11.5	2059	2	Q9PD50	xyella fas
3	966.5	9.3	2712	2	Q9F3X5	pasteurella
4	742	7.2	1299	2	Q9F3X6	pasteurella
5	704.5	6.8	1190	2	Q9PC04	xyella fas
6	690	6.7	1098	2	Q84152	haemophilus
7	677	6.5	1107	2	Q9F2D8	salmonella
8	658.5	6.4	3705	2	Q9F285	versinia pe
9	591.5	5.7	2340	2	Q9ZD91	rickettsia
10	559.5	5.4	2586	5	Q9VTK8	drosophila
11	557	5.4	2106	2	Q9XC47	rickettsia
12	553	5.3	2514	2	Q9JY30	neisseria m
13	540.5	5.2	4919	2	Q9ZHL0	haemophilus
14	528.5	5.1	2021	2	Q52657	rickettsia
15	526.5	5.1	2349	2	P94750	escherichia
16	526.5	5.1	2283	2	P76347	escherichia
17	526	5.1	3029	2	Q55582	synecocyst
18	513	5.0	4152	2	Q9ZHL3	haemophilus
19	512.5	4.9	2703	2	Q9K0T0	neisseria m

## ALIGNMENTS

RESULT	1	
ID	P71401	PRELIMINARY; PRT; 2353 AA.
AC	P71401;	
DT	01-FEB-1997	(TrEMBLrel. 02, Created)
DT	01-FEB-1997	(TrEMBLrel. 02, Last sequence update)
DT	01-JUN-2000	(TrEMBLrel. 14, Last annotation update)

OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision: Pasteurellaceae.

OC haemophilus.  
OX NCBI TaxID=727.

OX	NCBI_taxonomy=121;
RN	[1]

RP SEQUENCE FROM N.A.

RC STRAIN=C54:

RX MEDLINE=97047989; PubMed=8892830;

RA Geme J.W., Cutter D. III, Barenhamp S.J.;

RT "Characterization of the genetic l

RT. type b surface fibrils.";

RL J. Bacteriol. 178:6281-6287(1996).

DR EMBL; U41852; AAC44560.1; -.

SQ SEQUENCE 2353 AA; 243844 MW; 53E491B51A897B77 CRC64;

Query Match	11.6%;	Score 1206;	DB 2;	Length 2353;
Best Local Similarity	23.8%;	Pred. No. 1.1e-42;		
Matches 627;	Conservative 295;	Mismatches 853;	Indels 856;	

QY 1 MNHIYKVIFNKATGTTFMAVAEYAKSHSTGGSCATGQVGSVRTLSFARIAALA----- 53

1. \_\_\_\_\_  
 2. \_\_\_\_\_  
 3. \_\_\_\_\_  
 4. \_\_\_\_\_  
 5. \_\_\_\_\_  
 6. \_\_\_\_\_  
 7. \_\_\_\_\_  
 8. \_\_\_\_\_  
 9. \_\_\_\_\_  
 10. \_\_\_\_\_  
 11. \_\_\_\_\_  
 12. \_\_\_\_\_  
 13. \_\_\_\_\_  
 14. \_\_\_\_\_  
 15. \_\_\_\_\_  
 16. \_\_\_\_\_  
 17. \_\_\_\_\_  
 18. \_\_\_\_\_  
 19. \_\_\_\_\_  
 20. \_\_\_\_\_  
 21. \_\_\_\_\_  
 22. \_\_\_\_\_  
 23. \_\_\_\_\_  
 24. \_\_\_\_\_  
 25. \_\_\_\_\_  
 26. \_\_\_\_\_  
 27. \_\_\_\_\_  
 28. \_\_\_\_\_  
 29. \_\_\_\_\_  
 30. \_\_\_\_\_  
 31. \_\_\_\_\_  
 32. \_\_\_\_\_  
 33. \_\_\_\_\_  
 34. \_\_\_\_\_  
 35. \_\_\_\_\_  
 36. \_\_\_\_\_  
 37. \_\_\_\_\_  
 38. \_\_\_\_\_  
 39. \_\_\_\_\_  
 40. \_\_\_\_\_  
 41. \_\_\_\_\_  
 42. \_\_\_\_\_  
 43. \_\_\_\_\_  
 44. \_\_\_\_\_  
 45. \_\_\_\_\_  
 46. \_\_\_\_\_  
 47. \_\_\_\_\_  
 48. \_\_\_\_\_  
 49. \_\_\_\_\_  
 50. \_\_\_\_\_  
 51. \_\_\_\_\_  
 52. \_\_\_\_\_  
 53. \_\_\_\_\_  
 54. \_\_\_\_\_  
 55. \_\_\_\_\_  
 56. \_\_\_\_\_  
 57. \_\_\_\_\_  
 58. \_\_\_\_\_  
 59. \_\_\_\_\_  
 60. \_\_\_\_\_  
 61. \_\_\_\_\_  
 62. \_\_\_\_\_  
 63. \_\_\_\_\_  
 64. \_\_\_\_\_  
 65. \_\_\_\_\_  
 66. \_\_\_\_\_  
 67. \_\_\_\_\_  
 68. \_\_\_\_\_  
 69. \_\_\_\_\_  
 70. \_\_\_\_\_  
 71. \_\_\_\_\_  
 72. \_\_\_\_\_  
 73. \_\_\_\_\_  
 74. \_\_\_\_\_  
 75. \_\_\_\_\_  
 76. \_\_\_\_\_  
 77. \_\_\_\_\_  
 78. \_\_\_\_\_  
 79. \_\_\_\_\_  
 80. \_\_\_\_\_  
 81. \_\_\_\_\_  
 82. \_\_\_\_\_  
 83. \_\_\_\_\_  
 84. \_\_\_\_\_  
 85. \_\_\_\_\_  
 86. \_\_\_\_\_  
 87. \_\_\_\_\_  
 88. \_\_\_\_\_  
 89. \_\_\_\_\_  
 90. \_\_\_\_\_  
 91. \_\_\_\_\_  
 92. \_\_\_\_\_  
 93. \_\_\_\_\_  
 94. \_\_\_\_\_  
 95. \_\_\_\_\_  
 96. \_\_\_\_\_  
 97. \_\_\_\_\_  
 98. \_\_\_\_\_  
 99. \_\_\_\_\_  
 100. \_\_\_\_\_

Db 1 MNKIFNVNMTQTWVVSELTRTHK--RASATVETAVLATLLFATVQANATDEDEEL 58

Qy : 54 -----VLVIGA----- 59

59 DBW/BTA BVI SEUSDKFECTCEKEVTENSNCWCTVEFNKZCVI KACATYI KACONI KTWONED 11

DD JJ DFV VKIAFVLSFHS DREGIGEREK EVIENS NWGII FDNKGVLRAGAI I ENAGDNLKIKQNTU II

Db	179	DSTLPDVAINTGVLSSSFTPNDEKTRA	---	ATVKNLWAGNINCAKTAGNVSVDL	235
Qy	142	GGDVLAEGDASIAIGSDDL	---	YLPKNLNDKNEFHKLHGHILK	184
Db	236	---VSAYNNVEFITGDKNTLDVLITAKENGKTTEVKTPKTSVKEKDKLFTGKE	---		288
Qy	185	IOTSTDGKIKYRRRAOCHASTAVGAMSACQHPFNAAGTYATAEAYSIAVGLAQAATK	244		
Db	289	---NNDTNKVT	---	LVTAKAVID-AVNKAGRWK	328
Qy	245	QSSIAVGSNAKANAPAAFAIGNTVNLGRGVALGFGSQLDRDNNDUA	---	SAVPL	299
Db	329	TUT	---	ANGQNDPATVASCNTVFESGDGTAS	379
Qy	300	G	---	KTLDQYKATROGDSDFISGNSNNNSIRKIIINVAGSRODVAVNAQ	352
Db	380	GLKFSDDKIVADTTALTVTGK	---	VAETAKEDD	419
Qy	353	LKLVLELANRKTTFKGDGNN	---	SNSVERGLGNTLTIKGDAQTNALTEANLGVVTD	406
Db	420	-DLVTALGNLWKAKEADTDGALEGISKQOEKAGETVTFKAG	---	KNLKVQD	470
Qy	407	CNGLKVKLAKELTGLTSVATNKIIVTSNTNNNAELQSGGLTFSPITGKTDKTVYSIDG	466		
Db	471	GANTYSLQDALGLTS	---	ITLGGTN	508
Qy	467	LKFTNDSNIATGTRITRKKIGFAGTNDGVDESKPYL	---	DNEKLKVGNS	518
Db	509	LTITPAGNG	---	GTGTTNTISVTGDKAGNKAITNVASGLRAYDDANFDVNLNSATDLNRH	567
Qy	519	---	---	SGSLTVNNTGNKQIOVANGIKPATVANNVANTSAVTGARITEEKIGF	572
Db	568	VEDAYKGLLNLNEKANQPLV	---	TDSTAATVGLDR	610
Qy	573	DGVDEQAPYLDK	---	ERILKVGREITTDGGINAGNHITGLTNGIANT	622
Db	611	NGTKESNOVKQADEVLFTGAGAATVTSKENGKHTI	---	TVSVAETKADCGLEKDGDTI	667
Qy	623	KOLDAKPTLNAGDGIINSNGDLVDSSGNTTPTYNISVKTITKLSNGTSGNNKFSYS	682		
Db	668	KLKYDNQNT	---	DNVLTVGNGTAVTKGFE	712
Qy	683	NAHONNS	---	LVTAKDLADYLNKVNETADSLPSFKVONGDNS	736
Db	713	DATANDADKKVATVKDVATAIN	---	SAATFVKTEINTLTISIDEDNPTDNGKDALK	765
Qy	737	FNTLKLK-GEN	---	GVNIT	777
Db	766	GDULTFRAGKNLKVROGKNITFDLAKNLEVKPAKVSDTLTIGNTPTGGTTATPKVNIT	825		
Qy	778	SDTNGNRLVIEQVPSADGNSTKNT	---	IKGLSPTL	834
Db	826	STADGLNFAKE	---	TADASGSKNVLKGIATLTPESAGAKSSH	879
Qy	835	ASIDVVLNAGFNLKNGKDKDPVSTYIDVDIGNATATVYDEANQTSKVAYDVNWE	894		
Db	880	ASIEDVLRAGWNIQNGNNDVYATVDTVNFDTSDGTITVTQ	---	KADGKGADVKIGA	937
Qy	895	KTELTGDNKKOLGVKTKLTETSTNGNATTFSTDDH	---	ALVKASDI	941
Db	938	KTSVINKDHNGKLFYG-KDLK	---	DANNGATVSEDDCKDGTGLVTAKTVIDAVNKS	991
Qy	942	---	---	AGNLNTLAB	974
Db	992	RVTGEGATAETGATAVNAGNAETVTSQVSNFKNGNATTA	---	TVSKDNINIVKYDVN	1047
Qy	975	---	---	DK	1005
Db	1048	VGGLKIGDKKIVADTTITV	---	TGKVSVPAGANSVNNKLVNAEGLATALNLS	1103
Qy	1006	-DIKTOKDGVTVFGIINTQSGLRAGDSTTLN-NNGLSIRN	---	TASNEQIQVGADGVKE	1058

Db	1104	WTAKADKYADGESEGETDQEVKAGDKVTFKAGKNLKVQSEKDFYSLQDTLGLTSITL	1163						
Qy	1059	AMVNVVVGAGIDGTRITRDEIGFTGTNGS	---	LDKSKPH	---	LSKDGINAGKKITNI	1112		
Db	1164	GGTANGRNDTG	---	TVINKDGLITLANGAACTDASNGTISVTKDGISAGNKEITNV	1219				
Qy	1113	OSG	---	ELAKNSHDVATGGKIYDLKTELENKISSTAKTAONSL	1152				
Db	1220	KSALTKYKTQNTADETQDKBFHAAVKNANEVEFVK	---	EFVSADQGNFTVSNPYSSYDTSK	---	TSDVI	1183		
Qy	1153	H	---	BTVIDVAEKVGDGLEKDTGK	---	IKLKYDNTDGNLLTVDATKGVASVAKGFNAVTTDAT	1330		
Db	1271	HPVTTIDVAEKVGDGLEKDTGK	---	ITTKVANKGVVRVIGIDQTKGLT	---		1211		
Qy	1184	TPAGENG	---	ITTKVANKGVVRVIGIDQTKGLT	---		1211		
Db	1331	TAGTNNANERGVVVKGSNGATATETDKKVKATVGDVAKAINDAATFVKVENDSATIDD	1390						
Qy	1212	---	---	TPKLTVGNNG	1222				
Db	1391	SPTDDGANDALKAGDTLTLKAGKNLKVROGKNITFALANDLSVKSATVSDKLSLTNGN	1450						
Qy	1223	KGIVINSQGN	---	TITGLSNTLANVNDKGSVTTTGGNLIKDEKTR	1269				
Db	1451	KVNITSDTKGLNPAKDKSGTDDANIHLNGIASTLTDLLNSCAT	---	TNLLGNGIITONEKR	1509				
Qy	1270	AASIVDLVSGFNLQ	---	NGEAVDFVSTYDVFNFANGNTTAKVTVYD	---	DTSKT	1319		
Db	1510	AASVKYDLNAGNVRGPKASPANNQVENIDFVATYDVFVSGDKDTSVTVESKDNKR	1569						
Qy	1320	SKVYDVNVDDTTIEVKDKKLVKTTLTSTG	---	TGAN-KFALSNOAT	---	G	1365		
Db	1570	TEV	---	KIGAKTSVIKDHNGKLETKELKDANNNGVTVTETDCKDEG	1613				
Qy	1366	DALVKAADIWAHLNLTSGDITQFAGKASOANNSAGYVD	---	ADGNKVIVDSTDNKY	1418				
Db	1614	NGLVTAKAVIDAVNKAAGRVKTT	---	GANGQNDPATVASCNTVTFAGNGTTAEVT	1667				
Qy	1419	QAKNDG	---	TVDKTKEVAK	---	DKLVAQA	---	QTPDGTLAQMNKVSINKQVNDANK	1467
Db	1668	-KANDGSIYKVKVAVDGLKLDGDKIVADTTLTVADGRVTA	---		---	NGDGK	1716		
Qy	1468	KOGINEDNAFVKGLEKASDNKTKNAATVCDLNAVAQTPLTFAGDTGTTAKLGLT	1527						
Db	1717	K	---	FVDSAGLADALNKLWSA	---	TAGKEGTGEVDPAANSAGO	---	EVKAGDKVTF	1763
Qy	1528	KGGQDNTNKLTDNNIGVVAGTGTVKLAKDLTLNLSV	---	NAG	---	GTKIDEGIS	1578		
Db	1764	KAG	---	DNLIKIKOSGKDFYSLKKELKDLTSVEFKDANGGTGSESTKITKDKLT	1814				
Qy	1579	FVDANGQA	---	KANTPVLSANGLDGGKVISVNGKTK	---		---	DTDAAN	1619
Db	1815	ITPANGAGAAGANTANTISVTKDGISAGNKAATNVVGLKFGDGHTLANGTVADFEKH	1874						
Qy	1620	VOQLNEVRNLLGLGNDN	---	ADGNQVNIAD	---	IKKDPNSGS	---	SSN	1658
Db	1875	DNAYKDLTLNDEKADNPTVADNTAATVGLRGLGVISADKTTGEPNQEYNAQVRNAN	1934						
Qy	1659	RTVIRKAGT	---	VLGG	---	KGNNTTEKATGG	---	VOVG	1687
Db	1935	EVKFGSGINGINSGKTLNTRVITFELAKEGVYKSNFETVKNADGSETNLVKVGMGYSK	1994						
Qy	1688	VYKDGNGANDLSNVVKTOK	---	DGSKKALLATYNAAQOTNYVNNPAAID	1736				
Db	1995	EDIDPATSKPMTG	---	KTEKYKVENGVVANGSKTEVTLTNKSG	---	YVTGN	---	QVAD	2046
Qy	1737	RINEQGRIFRHHVNDGNOEPVQGRNGTDSASGKHSVAIGFOAK	---	ADGEA	1785				
Db	2047	AIKSGFEL	---	GLADAAEAEKAFESAKDKOLSKOKAETVNNHDKV	2090				
Qy	1786	AVAIGROTQAGNOSIAGDNAQATGQ	---	SIAGTGNVYVAGHKSAGIPSTVK	1837				
Db	2091	RFANGLNTRV	---	SAATVESTDANGDKVTTTFVKTDVLPLOLY	---	NTDANGNKIVKK	2144		





Qy	1009	TDKGTVTFTGINTQSGUKAGDSITLNNGLSIKNTASNEQIQVGADGVKFMVNGVCGA	1068
Db	999	NTDGNLTIS-----KSGSDNDV-----VFNLSKDFKVDGMTSGT--TVVNDGVKV	1042
Qy	1069	CID--GTRITRDEIGFTGNGSLDKSPHLKSDGINAGKKITNIQSGEIAKNSHDV	1125
Db	1043	GSDVALGTT-----GLTIANG-----PAVTASGIDAGSKVISHVAAGAVSETSDAV	1089
Qy	1126	TGGKIYDLKTELENISSATAQNSLHESFVADBOGNFTVSNPYSSDYTSKTSVDITF	1185
Db	1090	NGSQLNAVQASQPVFTG-----NEGAVKRSIGQSUVIS-----GESSTAGTY	1134
Qy	1186	AGENGITTVKVKNGVVRVIGIDQTLGTTPKL--TVGNNGRGIVINSQNGQNTITGLSNTL	1243
Db	1135	SGGN-LKSVVDGAAGRIHLQLA---DSPKFGNVVINGGK-----ISGVT---	1175
Qy	1244	ANVTNDKGSVRTTEOGNIIKDEKTRAASTIVDVLISAGFNLQNGEAVDFVSTYDVTNFA	1303
Db	1176	AGTEETDAV--NFSQLKSIATAVDQGWTLTASGANGSKVASGCTVDLKN	1222
Qy	1304	--GNTTAKVTYDDTSKTSKVVDVND-----DITIEVKDKLGVKTTTLTSTGTG	1353
Db	1223	TGKNLTISK-----SGDSNDVVFNLSKDFKVDGMTSGTTVVND---GVKVGSDVALGTT	1274
Qy	1354	ANKFALSNOATGDALVKAADIYVAHLNLTLSGDIQITAKGASQANNAGYVDADGNKVIYDST	1413
Db	1275	GLTIANGPAVTASGIDAGSKVISHV-----AAGAVSETSDAVNGSQLNAVQOAS	1325
Qy	1414	DNKYQAKNDGTVDKTEKAKDLKVAQAOTPDGTLAQMNVKSVINKE-----QVNDANK	1467
Db	1326	QPVTFGTG--NEGAVKRS--LQGSVVISGESSTAGTYSGGNLKSVVDDEAGTTHLQADSPK	1382
Qy	1468	KQGINEDNAPVKLEKAAASDNKTKNAAVTVGDLNAVQAOTFLTFAGDTGTAKKLGELITI	1527
Db	1383	FGNVVING--CKISGVTAGTEETDAVNFSQLKSI-----STAVDOGWTLTA	1427
Qy	1528	KG-----GQDTNKLNTDNNIGVAGTGDGTV--KLAKDLTNLNSVAGGTKIDKERGI	1577
Db	1428	SGANGSKVASGGTVDLKNTDGNLTISKSGDSNDVFNLSKDEK--KSIIVGNTQLDKGV	1486
Qy	1578	S-----FVD-----ANGQAKANTPV-----	1592
Db	1487	KVSSNVLDSNELVITSHSSTSSVKTLANGESVNVTVNGDGVNDDVVVNDLGLSTV	1546
Qy	1593	-----LSANGLDLGKVISNYKGTGKTDAAVQOLNEVRNLLGLG---NDNADGNQV--	1642
Db	1547	GGASLTLSGINAGSHKITNVTAGTEDTDAVNFSQLKSVSEAVDKGWTLTASGANGSKVYS	1606
Qy	1643	NIADIKK-DPN-----SGSSNRVTI-----KAGTVLGGKGNNDTEKLATGGVQVGVOK	1690
Db	1607	GGTVDLKNTDGNLAIKSGDSNDVFNLSKDFKVDYTAG-----NTVNTDGVKVGSDV	1661
Qy	1691	DCNANG-----DLSNVVTKTQDGSKKALLATYNAAGQTNVTVNPA---EADIRINEQGI	1743
Db	1662	SLGAMGLFIANGSPVTSAGFNAGOR---VISHVAVGADTDAVNVSQLKQAVQSVTVKAT	1718
Qy	1744	RFFHYNDGNOEPVQGRNGIDSSASGKHSVAIGFOAKADGAEEVAIGRQTOA--GNQSIAI	1802
Db	1719	RYIYSTNDGG---TOGGNYVDGATGSKAIAAGVGTQASGEGAAVSGAAASGKGSTAI	1774
Qy	1803	GDNAOATGDOOSTAIGTG-----NVVAGKHSG-----AIGDPSVTVKADNSYSVG	1845
Db	1775	GRNAIASADGSVALGDGAKDGGRGAEVSTGKYSGVQNNVTGTVSVGDAKAGETRSINVA	1834
Qy	1846	NNNOFTDAT-----QT-----DVEFGV--GNNTIVTRES---NSVAL	1875
Db	1835	DAKEAMDANVRLDVAQKSNLQTDMMRHETNNIEDVEFKITKGDSSASSVRGMGVNAMA	1894
Qy	1876	GSNSATSAGTHACTQAKKSDGTAGTTTATAGATGTYKGFAGGTAVGAVSVGASGAERRIQN	1935
Db	1895	GTNAAVS-GTESVALGKNTNVSNADNAVATG--NGSVADRA-----NSVSVGSGGSEBQVTN	1947
Qy	1936	VAAGEVSATSTDAVNGSOLYKATQSIANATNELDHRHIONENKANAGISSAMAMSPQA	1995
Db	1948	VAAG---TADTDVNVSQL---NQGLITAKQYTDGMVGNLRRRETSGVAAAIIATANLQA	2001
Qy	1996	YIPGRSMVTGGIATHNGOGAVAVGLSKLSDNGQWVFVKINGSADTOGHVGAAGVAGGFHF	2053
Db	2002	YVQGRGMTSVGVSSYQOGSAIAGVGSVAVSGHGVFKFSGSANTRSHVGVGAGVGYQM	2059
RESULT	3		
Q9F3X5		PRELIMINARY; PRT: 2712 AA.	
ID	Q9F3X5		
AC	Q9F3X5		
DT	01-MAR-2001 (TremBLrel. 16, Created)		
DT	01-MAR-2001 (TremBLrel. 16, Last sequence update)		
DT	01-MAR-2001 (TremBLrel. 16, Last annotation update)		
DE	MAPB PROTEIN.		
GN	MAPB.		
OS	Pasteurella multocida.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;		
OC	Pasteurella.		
OX	NCBI_TaxID=747;		
RN	[1] SEQUENCE FROM N.A.		
RP	STRAIN=PM70;		
RC	Henderson I.R., Nataro J.P., Cappello R., Stein C.;		
RA	"Evolutionary origins of the autotransporter proteins."		
RT	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
RL	EMBL: AJ277636; CAC14203.1;		
DR	EMBL: AJ277636; CAC14203.1;		
SQ	SEQUENCE 2712 AA; 276154 MW; 3F5579D6F32FFA3D CRC64;		
Query Match	9.3%; Score 966.5; DB 2; Length 2712;		
Best Local Similarity	22.0%; Pred. No. 1.6e-32;		
Matches	629; Conservative 338; Mismatches 932; Indels 959; Gaps 128;		
Qy	1	MNHLYKVIENKATGTTFMAVAYAKSH-----	26
Db	1	MNKYRVIVSWIVNTFTFAVSELATSKGVKFSAISNPQELNSSIPATFKLSAIALVS	60
Qy	27	-----STGGSCATQGVSVRTLSE-----ARIAALAVLVIG---	58
Db	61	ILAFAPSQVLAQDVNAVNLTVSGTSTFTGAATFNNTATFNNMVTAGNVATTLTVGKKNV	120
Qy	59	-----ATLNGSAYAQOI---TTKIEIGQTNKINNT-----	85
Db	121	ATKDDVTTLTTTVNQKAAQSEVDNLKNTKADKSDVDNLKNTKAAQSEVDNLKNTKADKSD	180
Qy	86	-----LKG-----	89
Db	181	VDNLKNIKADKSEVNEKNKADKSELTKLSNLVSLGVPGETTYYGKIFRVKSTLDD	240
Qy	90	ALATGEASIAFGSLSKAQGSQAIAIG-----SVKPDPNNGSNGNVS	131
Db	241	AVADQDSVAIGPKAKTEGDAVALGHDSHTANATESIAIGKKAYAVK--TANKGI--AIGE	297
Qy	132	HAK-----GNESIAIGDVLAEAGDAIAIGSDSDLYLPKLNLDL	168
Db	298	HARVGSKQDGVYVDGASSYVLGPKDGESSVAIGDKAVSRGEASIAIGN--AITSNKDA	355
Qy	169	KNEFKH-----LTHGHEI---LKKIQISTDQKIKY-----RRT	198
Db	356	KSOIAKNNTALGTNAQAIASDNSIAGNNAARTNKDSISIAIGDSAETKAAHSIAVGTTS	415
Qy	199	RAQGHASTAVGAMSTAOGHFSNAGFYATAEAYSLAVGLAAQATKQSSIAVGSNAKANA	258
Db	416	KALAEALALGKLAIEAKGTSSVAMGNTSKADGNSVAVGNTSQTLOSNTIAIGSSAIANP	475
Qy	259	FATATAGN-----TVNLSGR-----GVALGFGSQILDRDN---	289
Db	476	ERTISGLNAGKQEQADATGTRKHSQINIGENSGEVIGQLNIGIGHAGHTNVYGHNTIAL	535
Qy	290	-----NTD---ASAYVPLGKTLADQY-----KATROGDST-----	316

Db 536 GTYACTNLNSETSGANVSIGHE-ANKVDOLITAVOKSTVVGQATKAASRSTALGABATA 594  
QY 317 ---DIFSIG-----NSNN----- 326  
Db 595 LGLDAVAVGITSKAEQDKSVAICANSTADSNVALGATSRAVAKEGSGYLTKQSSILVVS 654  
QY 327 -----NNSIPRKLIINVCAGSRDIDAVNAQKLVLBEL----- 359  
Db 655 VHQDGAADQHILRLRVNVADGVDEQDAATVAQLKVKTEKVTESDLOAQFNAISHAPTA5E 714  
QY 360 -----ANKRITFK-----GDGDNNSVERGIGLNTLTIKGDAOTNALTE 398  
Db 715 IKYDTPPPPTCAENKIYILKAKTRISNVAPAELOTDVNLGOVNHIVTKNAHYFSVND 774  
QY 399 ANTGVVVT---DGNGLVKLAKELTGLTSVSATNKITVSTNNNN-----AE 441  
Db 775 TGINPVPNGHNDCAKALAMAIQONAAKABERSVAIGNNTTVNGEGSIGLGTYYKGSAR 834  
QY 442 LQSGGLTFSPITCTKTDKTV-YSID-CLAFETDNSNIATKGTTRITKKGIGFACITDNGVD 499  
Db 835 LQDGVPRESETTVTKPSKNVKGIAIGAGCTTTDGNNSIAIGSLAATSDK-----NPGAN 888  
QY 500 ESKPYLDNEKLVGNSGLTNSGSLTVNNTTGNKQIOVGANGIKFATVANNVANTSATVGT 559  
Db 889 VDR-----AIAIGYNAVSAEKA--NAIGDRAVANSVKGNAFGSALGSAESTAIOTE 940  
QY 560 RITEEKIGFA-----GTNDGVDEQ-----APYLDKE-RLK 588  
Db 941 SKSEGOVALGTPKSHAKGLNSTAFCTNQVVSQNSGSIYAGELGNAKATVINGEGTYS 1000  
QY 589 VGRVEIT---TDSGINAGNHKTGLTN---GIANT-----DAVTIKOLKD 627  
Db 1001 LGNTSTLTANESGIFGNSNEIKAKENARIVGNKNTIGAIEEKHPVPGTPPAAPVNDLK 1060  
QY 628 AKPTLNAGDGISNNSGDLVDSG-----NIT-TPTYNISVKTTLKNSGTSGNKK 678  
Db 1061 IYVT---GVDNKISSDKLAKULSGLFVYCHGNNTAQLDPDPTERTFTLTDSSVIGANNT 1117  
QY 679 FVSNAHDNNSLVTARDLADYL-NKVNETADSPALPKFQVQNGSNNAITVGRDITNGKTF 737  
Db 1118 L---NTKGRNYFVLGNVNTATLENSVYLGDASAYTTGNSTSSNNSYADMANGLNKSQYTF 1174  
QY 738 NTLKLGNGVNTTNRATGTVTFGIDQSNGLTTPKLTVCSDTNGNRL--VIEQVPSADG 795  
Db 1175 -----AGSOPVGVTVGVGKERRVQNVASGLVTEAST--DAINGSLFALTRPLRFAG 1227  
QY 796 NST---KNIIKGLSPILPSTASPSGRNIALG-----NTIEBKDKSNAASIDDLNAGF 845  
Db 1228 NSTLSNPCKPGDVTIVSRSSNOGKVVGGENDGNKLTITADKNIGVITANGDHTLEVR 1287  
QY 846 -----NLKN-----NGKDKDFVSTYDVTDFIDGNATTATVT---YDEANQTSKVAYDVN- 891  
Db 1288 AKTLSNLKDATFTGTDKTTINK-DGMTITNG-ANTVSLTEGGLNNGKNKITVAAGQNE 1345  
QY 892 ---VDEKTE-----LTGDNG---KQOLGVKTKTLTETSTNGNATFTSTDDHALVK 937  
Db 1346 TDADVNYKQNLNDLKAEGFGLTGEDGQTVKQALGT-AIKV--TGDDNKTIKVITDADGS--K 1400  
QY 938 ASDIA-GNLTALAEHTTKGTANTALQFTVKKVDENDKA--DDTNALITVKGDTSG-- 992  
Db 1401 KLEIGLENQVTLGGEAKGNPAADGKL---TLKNQAGTDKVVLDGANG-TVGLITGADGAQ 1456  
QY 993 KVNTLKLKGNGLD-----IKTRDKDGTVTVEG-----INTQSLKLAG 1028  
Db 1457 AVITVK-KGRPTLDNAETPRIAYGNEEVATLNDG-LKFEGANAGDVHNAKLTNQVDVKA 1514  
QY 1029 DSTTLANN---GLSIKNTASNEQIOVCADGVKPFAMVNGVVGAGI-----DGTTRIFR 1078  
Db 1515 TANTVWDFNFKGQNIIMTRVEGNTITVA-----LAKALSGLTSATFGDPASNPKDSTVINK 1569  
QY 1079 DEIGFTGTNCSLDKSKPHLSKQGINAGGKKITNIQSGELAKNS-----HDAV--TGGKI 1130

Db 1570 D--GLTITQGDNTVS---LTDGGLDNGNKQIKNVASGLTTTNGTATTSLDDAVQTNQVNV 1624  
QY 1131 YDLKTELENKISSTAKTAQNSLHEFSVADQGNFTVSNPYSDYTSKTSDDVITFAGENG 1190  
Db 1625 GDLKTAINNITNGT-----NPLGGFGLKDKAGNTF-----KQNLGETAQI--TGDSN 1669  
QY 1191 ITTKVANKGVVRCIDQTKGLTTPKLVGNNGKGVINSONGQNTITGLSNTLANVTNDK 1250  
Db 1670 VNTKV-----VD-----GONGKALEVSL---ANQLTLGRGPEANVPNAT 1706  
QY 1251 GSVRTTEOGNIIKDEKTRAASITVDLSAGFNLQGNCEAVDFVSTYDVTNPFANGNTTAK 1310  
Db 1707 G-----EACKITLKDKGTDRTVVVDGSEGAISLTGPATOGAAAAATAKIKVAGNPDLEN 1761  
QY 1311 VTYD---DTSKTSKVYVDV-NVDDTTI-----EVDKDKLGVK--- 1343  
Db 1762 TSDDPAMPNQNKTRITYDIAGPNGTITVTEQLATLNDGLKFGANTGVHDAKLNTRVDK 1821  
QY 1344 ---TTLTSTGTGANKFALSNOATGDALVAKSDIVAHLTLSGDIQTAKASQAN--- 1395  
Db 1822 GKAENTNANFDAGQ---IMTQISGNTIT---VALAKALAG-LDSATFGNPDGSKD 1872  
QY 1396 ---NSAGYVDADGNKVI---YDSTDNKYYQA-----KNDGTYDKTEV----- 1432  
Db 1873 GAVINKDGLTITEGDKTVKLTKEGLONGNQIINVDGSLKKTGDSVVALKDAEGSVLTNG 1932  
QY 1433 ---AKDLVAQAQTPDGTLAQNMVKSVINKEQVANDANKQ- 1469  
Db 1933 VNVGDLKNAIKDVTSATNGGFGGLKDKAGAEFKODLGTTAITGDKNINTKVIDVPNSDK 1992  
QY 1470 ---GIN-EDNAFY----- 1478  
Db 1993 ALBEISLANDITLKGNGADGVDSGLVNGKDGASVVLNGKDGSGITLTPRGQDGSKGKSAT 2052  
QY 1479 ---KGLEKASDNKTKNAVT-----VCDLNAVAQTPLTFAGDTG-TTAKLGE 1523  
Db 2053 ISVKDGKAGVDGDKGTTRIVYETKDATGKPVVEEATLNDGMKFVNDGKEVTRKLE 2112  
QY 1524 TLTIKG---OTDNTKLTDDNIG-----VVAGTDG--FTV 1553  
Db 2113 TLDIKGLDAATVADNAKVSNNLGVKTNAGTGLEIVMKERTFSLGLVNGKDGDAV 2172  
QY 1554 KLAKD-----LNLNSVNAGGKTKIDEK---GISFVDANGQAKANTPVLISANGLDL 1600  
Db 2173 KFAKDGKDGMSIAAVTD-NDGNATGLTIKDKDGNPGVTF---NNDGR----- 2215  
QY 1601 GGRVINSVNGKGTDAANVQOLNE-----VRNLLGLGNDNADGNQVNIADIKK 1649  
Db 2216 ---ITNVTAGVDKDAVNVSQLDKGLAKATTKVEAGKNMTVTPTVNDGSGTITYTVATED 2271  
QY 1650 DPNSSGSSNRPTVIKAGTVLGGKGNNDTEK-----LATGGVQGVYDK- 1690  
Db 2272 NVNF-----TTVTITGNTVM---NNDGVKVDNVALTNEGLKAGDVTVTTAGINAGNKV 2322  
QY 1691 DGNANGDLS-----NVWVKQKDGSKKALLATYNAACQNTVYNNPAEADIRINEQG 1742  
Db 2323 TGVAADGISPNSTDAVNGSQLNAVKETAEGAWHLTANGADSSNV---KPRNTVDLNNITDG 2379  
QY 1743 -IRFFHVDGNQBPVVGVRNGRIDSSAGSKHVAIGFOAK--ADGEAAVAIGRQTA-QNQ 1798  
Db 2380 NIVISKTNTAKINVT---FGLADNINVKDSVVVGPKANGKPGEGAVVINAEDGANGKD 2436  
QY 1799 SIAIGDNAOATGDSIAI-----GTGNV-----VAGKHSQA-IGDPSTVKAODNSVGN 1846  
Db 2437 GISI---VGKDKRDAVAISGDKGVGTIGLTGPAGADGKNANAIIGVNDSVKGLDNGDKD 2493  
QY 1847 NNOFTDATQTDVFGVGNNTITVTSNVSALGNSAISAGTHAGTOAK-KSDGTAGTTTAT 1905  
Db 2494 GNSKTRIVTKPNEEEQV-ATMNDGLVFGADGTEHAKLGTITVKVKGDDKNIETEVAG 2552  
QY 1906 ATCTVKGFAGQTAAGVAVSGAS-----GAERRION-----VAAGEYSATSDAVNGSOLY 1955  
Db 2553 DTLRVR-LKDNIDVKGINVTENUTVKEGAKINMGNVNDGVADGEVNAATSKOAVNGSQLH 2611

Thu Sep 13 14:19:33 2001

QY	1956	KATOSI---ANATNELDRIHONENKANAGITSSAMAMASPOAYIPGRSMVTGGIATHNG	2012
DB	2612	KVOOVNQAIAINKLGDHINKVDKDLRAGLAGATAVAFLOPNEAGKSIVSLGVGSYRS	2671
QY	2013	OGAVAGLSKLSRDNOWFKINGSADTQGHV--GAAVG	2048
DB	2672	ESAIAGVYARNSDNNKISIKLGGGMSRGVDNFGGSIG	2709
RESULT	4		
Q9F3X6			
ID	Q9F3X6	PRELIMINARY; PRT; 1299 AA.	
AC	Q9F3X6;		
DT	01-MAR-2001 (TremBrel. 16, Created)		
DT	01-MAR-2001 (TremBrel. 16, Last sequence update)		
DT	01-MAR-2001 (TremBrel. 16, Last annotation update)		
DE	MAPA PROTEIN.		
GN	MAPA.		
OS	Pasteurella multocida.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;		
OC	Pasteurella.		
OX	NCBI_TaxID=747;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PM70;		
RA	Henderson I.R., Nataro J.P., Cappello R., Stein C.;		
RT	"Evolutionary origins of the autotransporter proteins.;"		
RL	Submitted (APR-2000) to the EMBL/GenBank/DOBJ databases.		
DR	EMBL: AJ277635; CAC14202.1; -		
DR	SEQUENCE 1299 AA; 130963 MW; 8BCCE0EB66CDB428 CRC64;		
QY	1	MNHYKVFIEFKATGTFMAVEYAKSHSTGGSCATQQ---VGSVRTLSFARIAALVVI	57
DB	1	MNKYIFLWNAATQSVVWSELAKA---GGKSASKSALVNSVGSFSTLIAASVVL--	54
QY	58	GATLNGSAYAAQIITKEIGTKNTKINTLKGDALATGEASIAFGSLSKAQSGAIAIGSV	117
DB	55	---GS---GOVNAAEVT-----	65
QY	118	KPDPMNGSNGVSHAKGNESIAIGDVLAEQDASTAIGSDDLPLPKNLDLKNFPHKLIH	177
DB	66	---GNTG---VSGDDKYCFYNASSQS---VIC	88
QY	178	GHEILKKTQTDKIKYRTRFRAQGHASTAVGAMSYAQGHFSNAGFYATAEAYSILAVG	237
DB	89	GDATTK---ITD---KTDNPKAKSVVIG-----FG---ATNDGETNVAIG	124
QY	238	LAQAOTKOSSTIAGVSNKANAFATAIGGTVVNLGRGVALFGSQILDRDNNNTDASAYV	297
DB	125	AKSKSKAASIAIGNAKALDNOAIAIQNATANSDDWDISIG-----	166
QY	298	PLGKTLADQYKATRO--GDSTDIIFSIGNNNNSIRKIIINWAGSRDQDAVNAQIKL	355
DB	167	---RQGAEEQTEVSAEGRN---IAIGDGA-----	191
QY	356	VEELANRKITFKGDGDNNSNVERGLGNTLTIKGDAQTNALTEANIGVWTDGNGIKLVKLA	415
DB	192	---RGKGVNNIAILGTAGDRLACTHNLGMGTVNADEAV-----	228
QY	416	KELTGLTSVSATNKIIVSNNTNNNAELQSGGLTFSPITGTKTDTKYSIDGLKNTDNSN	475
DB	229	--RSALTAGSTTKKEINAKETDNK-----YYIEA-----SNT	258
QY	476	IATKGTTRTKKIKIGFAGTNDGVDSEKPYLDNEKLKVGNTLSNLSGLTV-----NNTTGN	530
DB	259	VAL--GTRALATQLAAYVA-----IGQAKAFGNGQSVAVGNGTKASQTAIATGNSNAHATGS	312
QY	531	KOIOVGANGIKFATVANNVANTSATVGTARITEEKIGFAGTNDGVDQEQAPYLKERLKVG	590
DB	313	SSIAIG-----GIVNGN--DTARTLASDNFTIAM-----GLSAQATKSD--AIAVG	354
QY	591	R---VEITTOS--GINAGNHKIIITGNTGIAN-----TDAVTIKQLKDAKPTLNAGDIS	639
DB	355	RNAKAAGINTVSIYAGVTKTADAQSDPNKLEPATDAVFIGNKAGYKSNQRMQ--VS	413
QY	640	INSNNGDLVDSGNITPTTYNISLVKTKIINSNGTSGNNKFSVNAHDN--NSLVTAQDL--	696
DB	414	LCKDSGEGVGTENTVIG--NSAGKNTKGTNTVAISSRACQNVGECHDNFAALIEAGONIK	471
QY	697	ADYLNKVNEDASALPSFKYONGDNNNAITVGTDKTNGKTFNTLKLKGNGVNTTNRA	755
DB	472	GSDNIAIGHAGRSADPNTKL---NINNTISLKE-----SVSLA-----	508
QY	756	TGTVTFGIDQSNGLTTPKLTGSDTNGNRLVIEQVPSADGNSTKNIKLSPTLPSTASP	815
DB	509	---NFGIAQGNKAKTDGLA-----SIAIGRNAEAVGGGTETANIAIGDSAS---ADA	552
QY	816	SGRNIALGNTIEEKD-----KSNASIDDDVLNAGFNKLNKNGKDKDFYST	859
DB	553	SGA--IVLGTAKAOKSLTVDGKKYGAYSAIVIGTEAKATAIAAAPAGKN--ENPKDAIAGT	609
QY	860	YDVFIDGNATATVTYDEANQTSKVAYDVNVDEKTIETLTDGNGKKQLGVKTIKL--TET	918
DB	610	KAPAHY-----ASTIALGFGAKSDTKAQAVSIGYN-----SNAK---GYQAIAGFSEA	654
QY	919	STNGNATTFSTDDDDHALVKASDIAGNLNTLAEIHTTKGTANTALOTPTVKKVVDENDKAD	978
DB	555	KTENA-----GSSIA--FGTKAQ---TRASASIAIGMGAETGDFDGOALD	595
QY	979	DTNATVKGDTSGKVNLTCLKGNGLDIKTDDKGTVTFTGINTOSGLKAGDSTTLNNNGL	1038
DB	696	GSDAVALGREA-----KAKRONAL-----AFGYKAVADHK-----	725
QY	1039	SIKTASNEQIQVGADGKPFAMVNGVVGAGIDGTTTRIDREIGFTGNGSLDKSKPHLS	1098
DB	726	---DAVALGA-----GAETAAEGTNEATVNEEFKYSGFAG---IKPIAT	763
QY	1099	KDGINAGCK---KITNTQSGEIAKNSHDVATGKIVYDLKTELENKISSHTAKTAQNSLHE	1154
DB	764	---VSVGKDAERTINVAGRIDKSTDAINGSQLY-----LALNAL--	803
QY	1155	FSVADEQGNFTVSNPYSSYDTSKTSDVITFAGENGITTKVKNKGVVRVGDITQKGLTTPK	1214
DB	804	---AG-----	829
QY	1215	LTVGNNGKGVINSQNGQNTITGLSNTLANVTDKGSVRTTEQGNIIKDEDKTRAASIV	1274
DB	804	---GN-----VGNTL--VTNVLGG-----DAAIVKEGDE-----	827
QY	1275	DVLSAGFNLQNGEAVDEFVSTVTVNFANGNTTTAKVTYDDTSKTSKVVDVNVDDTTIE	1334
DB	828	---AG-----	829
QY	1335	VKDKKLGKVVTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNTLSGDIQTAKGASQA	1394
DB	830	---TLTWSNIGG-----TGKG-----	842
QY	1395	NNSAGYVDADGNKVIYDSTDNKKYQAKNDGTVDKTKEVAKDKLVAQAQTPDGTLMQNVK	1454
DB	843	---TIHDA-----IE	849
QY	1455	SVINKQVNDANKKQGINEDNAFVKGLEKAAADNKTNAAVTVGDLNVAQOTPLTFAGDT	1514
DB	850	AVNNTAKASKTTVKEG---DNITV--TEEAAD-----	877
QY	1515	GTTAKKLGETLTIKGGQTDNTKLTDDNIGVAGTGTGFTVKLARDLTNLSNVAGTKTIDE	1574
DB	878	---GSRVTYVATKKQV--KFDVSVVAGGTKIDA	904
QY	1575	KGISFVDANGQAKANTPVLSANGDLGGKVISNVGKG--TKDT--DAANVQOL----NEVR	1627

Db 905 NGTFVDDQTKDNTFSPISKTDGIDAGNQVTVNQNGIAKDSKDAVNGQLFAQEGVK 964  
QY 1628 NLGLGNDNDAGNOVNADIKKDPNPGSSNRVTIRKAGTVLGGKGNNDTEKLTAGGVQVG 1687  
Db 965 NIIG-----GDTFNPETGEYAN-----TWIGTGASTI----- 993  
QY 1688 VDKDNGNGLDLSNVWTKQDGSKKALLATATYNAAGQNTYNNPAAIDRINEGGIRFFH 1747  
Db 994 -----DEAKAVNTTAKAA----- 1007  
QY 1748 VNDGNOBPVVOGRNGIDSSAGKSHSVAIGFOAKADGAAVAIGRQTOAGNQSIAGDNGAO 1807  
Db 1008 -----KTEVVOGENIVVTSAPGN-----GNTVTVATAKE 1038  
QY 1808 ATQDQSTAICTGNVNAKKGSAIGDPS--TVKADNSYSVGN--NNQFTDATQTDVFGV-- 1861  
Db 1039 VTFDKTT---VGSVWTDKNTDITGLSKNTLGGDNFAKNRAASEEQLNATQTNLATLLG 1095  
QY 1862 GN-----NITVTESNSVALGS--NSAISAGTHAGTQAKK-----SDGTAGTTTAGATGT 1909  
Db 1096 GNAQNTNGNVAMTDIGTGGKNNINDAIKASRNEVKOGKNNMVVPTTTCANGQTIYEATAD 1155  
QY 1910 VKGFAGOTAVGAVSVGASGASGAEIRIONVAAGEVSATSTDAVNGSOLYKATOSIANA---TN 1966  
Db 1156 KVAF-DEVKVGGITIDAT--TNKISGIAKGDISENSTDAVNGSOLYELQOKIAKSGDNYN 1212  
QY 1967 ELDRHQNENKANAGITSAMAMASPOAYIPGRSMVTGGIATHNGQGAVAVGLSKLSDN 2026  
Db 1213 ILNRRINKVBDLRAGIAGAGANAAGLPOAYIPGKSMVAVAGTYKGNALGMSRISDN 1272  
QY 2027 QQWVFKINGSADTQGHVGAAGAGFHF 2053  
Db 1273 GKVIILKTGNTNSRGDPFGASIGAGYOW 1299  
  
RESULT 5  
O9PC04 PRELIMINARY; PRT; 1190 AA.  
AC G9PC04;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE SURFACE PROTEIN.  
GN Xf1981.  
OS Xylella fastidiosa.  
OC bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
OC Xylella.  
OX NCBI\_TaxID=2371;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9A5C;  
RX MEDLINE=20365717; PubMed=10910347;  
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
RA Alvarenza R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,  
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
RA Ho P.L., Hohlseil J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
RA Krieger J.E., Kuramae E.B., Laigret F., Lambais M.R., Leite L.C.C.,  
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
RA Marquês M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,

RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
RA Zago M.A., Zatz M., Weidman J., Setubal J.C.,  
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";  
RL Nature 406:151-159(2000).  
DR EMBL; AE004017; AAP84783.1; -;  
SQ SEQUENCE 1190 AA; 118446 MW; 756741B0C8D787CC CRC64;  
  
Query Match 6.8%; Score 704.5; DB 2; Length 1190;  
Best Local Similarity 22.9%; Pred. No. 5.7e-22;  
Matches 360; Conservative 179; Mismatches 508; Indels 525; Gaps 65;  
  
QY 624 QLKDAKPTLAAGDGSINSNNGDLVDSSGNTTPTNYI-----SVKTKLNSGT 673  
Db 2 QIHTALPMVRLGG-----IEGVDPFLPKYKIGSLQHAVMTSAASSKKGTPR 50  
QY 674 SGNNKFSVNA-----HDNNSL--VTAKDLA---DYLKNKVNETAADSPKFKVQNGDSNNA 725  
Db 51 RSNNAMTAKRSAKLHRRQLHIVLLTVLAASTGYTGKAAQV-----VYNSDSTENC 102  
QY 726 ITVGKDTNGKTF-----NTLKLKGVGN-----ITTNRA 755  
Db 103 VETLGSSQTSFTHSASNDCKPDTQTEYSLEFYDRLNVLGSLYNEGKGLVDISGA 162  
QY 756 TGTWTFG-IDQSNGLT--TPKLTGVSDTNGRLNVEQVPSADGN--STKNILKGLSPTLPSI 812  
Db 163 TYSWRLGSIATMGSAGIDSIAGSG-----QSKTDTGNTSGATVAQGL----- 206  
QY 813 ASPSGRNIALGNTIEEKDKSNAASIDDLNAGFNKNGKDKDFVSTYDVFIDGNATT 872  
Db 207 -----RSIAIGTTARSQSQ-----DAISIGTCATG--NFAIA-----IGNALT 245  
QY 873 ATVTYDEANQTSKAVDVNVDEKTIETLNGKKQLGKVIKLTETENG-----NATT 926  
Db 246 SIANGIALGASSVT-----TRGGVALGQGSAAATASGITGYDPYTKSTST 291  
QY 927 FSTDDHVALKASDIAGNLNLTABEIHHTTKGTANTALQTFVKKVDENDKADDDNATVVG 986  
Db 292 LSTSMWRSTLGAVSI--GNITSSSTQRLTGLA-----AGRSDFDANVA 335  
QY 987 KDCSTCKVNTLKLKGN---GLDIKDKDGTVTTFGINTQSLKAGDSTLNN-----NGLSI 1040  
Db 336 Q-----LKLAEVSGGWNLTAS-----GANS--SNVALGESVDLKNSDGNLLIT 378  
QY 1041 KNTASNEQIOVGADGVKFAVNVNNGVCGAGIDGTT---RITRDEIGFTGTNGSLDKSPHL 1097  
Db 379 KTTDSNDVTFNLTALFKVDSLTTGNTAMTTDGVTVGKRVTLTDLSTGLVIAEG-----PSV 432  
QY 1098 SKGGINAGGKKITNIOSGEIAKNSHDAVGTGKIYDLKTELENKISSTAKTAQNSLREFSV 1157  
Db 433 ISSGINAAGKIMNVGTGA---DTDAVNGQL-----QAVSDTA----- 469  
QY 1158 ADEQGNFTVSNPYSSYDTSKTSVITFAGENGITTKVKNRGVVRGIDQTKGLTTPKLTV 1217  
Db 470 ----- 469  
QY 1218 GNNNGKIVINSQNGQNTITGLSNTLANVTNDKGSVTTTQGNIIKDEKTKRAASIVDVL 1277  
Db 470 --SKGMNLLASGNTSSNVAPCASVDLAKNT-----DGNLL-----ITKAIGINDVT 512  
QY 1278 SAGFNLOGNCEAVDFVSTYDTVNFANGNTTAKVTYDTSKTSKVYVDVNVDDTTIEVKD 1337  
Db 513 ---FNLATALEA--DSLTTGNTAMTTDGVTVGSNVT---LSTGLVIITD----- 553  
QY 1338 KILGVKTTTTLTSTCTGANKFALSQAQTGDA---LVKASDVAHLNLTLS--GDIQTAKGASQ 1393  
Db 554 -----GPSVTSISAGNKIITNVAAGTADTDVAVNFSQLQVAVSSSTASKGNLLASCANS 607  
QY 1394 AN-----NSAGYVDADGNKVIYDSTDNKKYQKNGDGTVDKTKVAKDLVA--QAQTPDGT 1447

DR EMBL: U38617; AAC43721.1; -.  
SQ SEQUENCE 1098 AA; 114100 MW; D977335A89F7333D CRC64;  
Query Match 6.7%; Score 690; DB 2; Length 1098;  
Best Local Similarity 24.7%; Pred. No. 2.1e-21;  
Matches 314; Conservative 151; Mismatches 439; Indels 366; Gaps 48;  
QY 953 HTTKGTANTAA---LOTETVKKVDENDKADDTNAITVGGDGTSGKYNT----- 996  
DB 26 HTKCASTAVAVAVLATLLSATVEANNPTVINKLAKYGDANFNFTNSIADAEEKQVQBEAY 85  
QY 997 ---LKLKGNGLDITDKDGTVTFTGINTQSGLKAGDSTTLNNG--LSIKKTASNEQIQV 1051  
DB 86 KGLLNLNEKNASDKLLVEDNT-----AATVGNLRLKLGWLWLSKKNGTRKESQ- 132  
QY 1052 GADGVFAMVNNVCVAGIDGTRITRTRDEIGFTGTNGSLDKSPHLSKDGINAGCKKTTN 1111  
DB 133 ---QVXHA-----DEVLEFEG-----KGVQVQTS 152  
QY 1112 IQSGEIAKNSHDAVTCGKIYDLKTELENKLSSTAKTAQNSLH--EFSVADEOGNFTVSN 1169  
DB 153 -----TSNGKHTITFALAKDLG----- 170  
QY 1170 PYSYDTSKTSVDITFAG--ENGITT--KVNKGVRVVGIDQTKGLTTPKLTVGNNGKG 1225  
DB 171 ---VKTATVSDTLTIGGGAAGATTTPKVV-----TSTDLGLKFAKDAAGAGDITTV 220  
QY 1226 VINSQONTITGLSNTLANVTNDKGSVTFTEOGNIIKQBDK--THAASIVDVLSAGFNL 1283  
DB 221 HLN-----GIGSTLTDLT--VGSPTHIDGG---DQSTHYTHAASIKDVLNAGWNI 266  
QY 1284 Q-----GNGEAVDFVSTYDTVNFANGNTTAKVTYDDTSKTSKVVDVNVDDTTIEV 1335  
DB 267 KGVKAGSTTGOSENVDVFTVDFEFLSADTETTTVTVDSEKENGKTEVKIGAKTSVIRE 336  
QY 1336 KDKKLGV-KITTLTSTGTGANKFALSNOATGDALVKRASDVAHLNLTSGDIQTAKASQA 1394  
DB 327 KDKLFTCKANKETNKVDGAN--ATEDADEGKGLVTAKVIDAVNKTGWRIKTWDANGON 384  
QY 1395 NNSAGYVD-----ADGN--KVIYDSTYDNKYQAKNDGTVDKTKEVAKDLVAQAQTPD 1445  
DB 385 GDEATVASGTVTFASNGTATTATVNTGTDG--ITVKYDAKVGDLKLDGDKIAA----D 437  
QY 1446 GTLAQMNVKSVINKEQVNDANKKOGINEDNAFVKLEKAAASDNK---TKNAAVTVGDLNA 1502  
DB 438 TTALTIVN-----DGKANNPKGVADVASTDEKKLVTAKGLVTA--LNS 479  
QY 1503 VAOTPLTFAGDTGT-----TAKKLGETLTIKGGQIDTNNKLTNNIGVVGAGTGFVKL 1555  
DB 480 LSWTTTAAEADGGTLDGNASEQEVKAGDKVTFKAG-----NLKVKQEGANFTYSL 530  
QY 1556 AKDLTNLSV-----NAGGTRKIDEKIGSPVDANGQA--KANTPVLSSANGLDLGGKVISN 1607  
DB 531 QDALTGLSITLGTGNGAKTEINKDGLTITPANGAGANNANTISVTKDGISAGSQSVKN 590  
QY 1608 VGKGTKDTDAANVOQLNEVRNLLGLGNADGNQVNIADIKDPNCSGSSNRTYKAGTV 1667  
DB 591 VVSGLKFKGDANFDPLTSSADNLTQNDODAYKGLTNLDEKGTQKQTPVVAQNTAATVGD- 649  
QY 1668 LGKGNNNDTEKLATGQVGVVDKNGDNLNVMVVKQDKGSKKALLATYNAAGQTYNV 1727  
DB 650 LRGLGWNVISADKTTGSGTEYHDOVRNNE-----VKF-KSGN-----GINVSGKT--- 693  
QY 1728 TNNPABEADIRNEQGIREFPHVNDG-----NORPVVOGRNGIDSSASGKHSVAIG--FOAKA 1781  
DB 694 -----VNGRREITFELAKGEVVKNSNEFTVKETNGKETS-----VKVGDYKYSKE 738  
QY 1782 D-----CEAAVAIGROTA-----GNOSTAIGDNAQATQDSIAIGTGNVNAKHSAGIG 1831  
DB 739 DIDLTGQPKLKDNTVAAYQDKGKVVSVTDTEATITNK---GSGVYTCNQVADAI- 794  
QY 1832 DPSTVKADNSYSYGVNNQ-----FTDATQTDVFGVGNNTVTESNVALGSNSAISACT 1885

DB 608 SNVYPGSDVLKNSDGNLLITKTDS-----NDVTENLATALKVDLSLTGTNTAMTDTGV 661  
QY 1448 LAQMNVKSVINKEQVNDANK--KOGINEDNAFVKLEKAAASDNKTKNAAVTVGDLNAVAQ 1505  
DB 662 TVGSNVTLGSLVTDGSPVTSSEGI-----SAGNQKITVNAAGTADTDAVNF 709  
QY 1506 TPLTFAGDTGTAKKL-----CETLITIKGGQDTNKLTDNNIGV--VAGTDGTFVTKAKDL 1559  
DB 710 SOLQAVSSTASKGNLNLASGANSNVAFGESVDLKNKDGNIVISKESGSDNVLFLNSSL 769  
QY 1560 TNLNSVNAAGTKIDBKGTSE-----VDANGAKANTPVLSSANGLDLGGKVLNNGKGTGD 1614  
DB 770 -KLDKLTVDGTVMTNGTVGSGVTLGSMGLVITDGPVSTSSGINAGSQKITVNAAGTAD 828  
QY 1615 TDAANVQOLNEVRNLLGLGNDNADGNQVNIADIKDPNCSGSSNRTYKAGTVLGGKNN 1674  
DB 829 TDAVNLSQLN-----TAMAGSGAK 847  
QY 1675 DTEKLTAT--GGVQVGVVDKNGANDLSNVWKTQDKGSKKALLATYNAAGQTYNVTNPA 1732  
DB 848 SVHYSTYDGGTQ-----GCNTYNGD----- 867  
QY 1733 EADIRNEQGIREFPHVNDGNQEPVVOGRNGIDSSAGKHSVAIGFQAKADGEAAVAIGRQ 1792  
DB 868 -----GATGTRSIAGVGVGTLSAEGATAVGSG 894  
QY 1793 TQA-GNQSTAIGDNAQATQDSIAIGT-----NVVAGKHS-----AIGDPS 1834  
DB 895 AASGKGSTAIRNAVASADGVALDGAKDARGAESYTGKYSGLQNNVTNVTGVSVDAS 954  
QY 1835 TVRADNSYSYGVNNQTFDAT--QTDVFG-----VGNNI-----TVTESNSVALGSN 1878  
DB 955 KGETRTVSNVADAKEATDAVNLQDLDRVAQDANRYVDNKIESLSEGOFPVKVNSL--NN 1011  
QY 1879 SA--ISAGTHA-----GTOAKKSDGTA-GTTTTAGATGTVKGFAGQTAGV-----A 1921  
DB 1012 SATPIAGVDATAIGVGATAGSADGADSIAMGNKASASADNAV-----AIGNHSVADRAWT 1064  
QY 1922 VSGCASAERRIQVNAAGEVSATSTDAVNGSOLYKATQSTANATNELDHRIHQENKANA 1981  
DB 1065 VSGSAGSERQVTVNAAG---TADTDVNVSQL---NOGLITAKQYTDGVVGLSRDRTDG 1118  
QY 1982 GISSAMAMASPOAYIPGRSVNTGQGIATHNGQGAVALGSLKSDNGQWFKINGSADTQ 2041  
DB 1119 GVAAIATANLPQAYIPGRGWTSGVSSYRGQSAIAVGSSVESGRWVFKFSGSANTRS 1178  
QY 2042 HVGAAVAGAGTFH 2053  
DB 1179 QVGIGAGVGYQW 1190  
RESULT 6  
Q48152 ID Q48152 PRELIMINARY; PRT: 1098 AA.  
AC Q48152;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE ADHESIN (HIA).  
GN HIA.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NONTYPEABLE STRAIN 11;  
RX MEDLINE=96332658; PubMed=8730864;  
RA Barenkamp S.J., St Geme J.W. III;  
RT "Identification of a second family of high-molecular-weight adhesion  
RT proteins expressed by non-typable Haemophilus influenzae.";  
RL Mol. Microbiol. 19:1215-1223(1996).



Thu Sep 13 14:19:33 2001

01-MAR-2001 (TReMBLrel. 16, Last annotation update)

DE YAPH. PROTEIN.  
GN Yersinia pestis.  
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Yersinia.  
OX NCBI\_TaxID=632;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=CO-92 BIOVAR ORIENTALIS;  
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;  
RT "Evolutionary origins of the autotransporter proteins."  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ277631; CAC14227.1;  
SQ SEQUENCE 3705 AA; 370756 MW; 714PDF16455968C9 CRC64;

Query Match 6.4%; Score 658.5; DB 2; Length 3705;  
Best Local Similarity 21.8%; Pred. No. 2.3e-19;  
Matches 523; Conservative 325; Mismatches 925; Indels 631; Gaps 118;

QY 1 MNHLYKVFENKATGTFMAVAEYAKSH----- 26  
DB 1 MNIFKVIWNASLVNVVVVSELAKGRITKSSRNLSISEGLPKFQSMVSKLFRKNLLAL 60  
QY 27 -----STGGSCATQGVSVRTLSFARIAALA-----VLVIGA--TLNGSAVAQOIT 71  
DB 61 SLGSIVFLSTGPGFAADITVSTQAELS-----AALSNGTVYDKILGADITLIGSLVNMTS 116  
QY 72 TKIEIGOTNK-----INNLTG-GDALATGEASIAFGLSIAQGSQAIAIGSVKPDNPNGSN 126  
DB 117 NQVVIDGGFGLTVNNTWYGLVSSGGTTLTQNNKIDSANYVMVL-----NGAN 171  
QY 127 GNV-----GSHAKNESIAIGDVLAEAGDASIAIGSDDLVLPKNLDL 168  
DB 172 TAVNVIYNNIDFLGSSQLIYMGAYGAATNSIMTFGDIL--NDVVNDRAQGEYVKNLAF 229  
QY 169 KNEPHKIHGEILKKIQT-----STDGKIKYRRTAOGH--ASTAVGAMSYA-- 214  
DB 230 TGRFHVHTGSSVTSFSTGCGANTSTMDPASGADVKIDRTGTGDLTSTGVNAFAYTEA 289  
QY 215 -----QCHFSNA-----FGYATAEAYSLAVGLAAQATKQSSIAVGS--NA 254  
DB 290 DGASPELLANQVFSGTTNRGLEIGSYNSID-FGSGVKIVLOSRSDCGLIISGNGIDNA 348  
QY 255 KANAFAA--TAIG-GNTVNLGRVAL-FGGSQILDRDNDNTDASAYVPLGKTLADQYKAT 310  
DB 349 TTNAAGINNAGDANVIYNLGTGSLKATWTGILATKNANAS-----DIY-IR 397  
QY 311 ROGSDTDIFSGNSNNSSIRKIIIVGAGSRDIDAVNVAQLKVE-ELANRKITPKGD 369  
DB 398 SAGDITATGATGISAHTGTGV--KIKNDGTTTSTAGIAISSAIKELISVDNTDGTITAT 455  
QY 370 GDNNSVVERGLNLT--TIKGAOTNALTANIGVVDGNGLKVAKELTGLTYSVAT 427  
DB 456 AGTVGNVLASALNLEFGGTINTSATANGITAG-----TEGCHTLTDLTINLLG-TGIALS 510  
QY 428 NKITVSTNNNAELQSGGLTSPITGKTKTKTVYSIDGLKFTDSDNSIATKGTTRTKK 487  
DB 511 NVAGVNLTLN-----VTLNTLNGT-----ALNSLTGLTLVDSLNG--RNTINIEGA 555  
QY 488 KIGFAGTNDGVDESKPYLDNEKLK-----GNSPLNSGSLTVN----- 525  
DB 556 GIGIAATNTLNT-----FPAEALDINVAGCIGIATGGVNLASNLIIINVANTLGTAL 611  
QY 526 -----NIT--GNKOIQVANGIKPATVANNVANTSVATGRTARTEKIGFAGTNDGVD 576  
DB 612 QITDGDINTTTIGN-ETQLNAEN--ATAINFLGSSKTLNNNGTIKGSVIFAGVAD--- 664  
QY 577 EQAPYLDKELKVRVEITDSCINAGNIKITGLTNGIANTDAVTIKQKDAKPTLNAGD 636  
DB 665 -----HIINNNTLDGT-----LITGAGN-DTLVLIDSSSQSNDVINLGD 702

QY 637 G-ISINNNGLDVS-----SGNITPTTYNISVKTTKLNS-NGTSGNNKFSVSNADHNNSL 690  
DB 703 GNSVTIQNGATVSSIIITGNGNDFTINGMSVSGTSLGSLDAGTGLATXNXXASTD--EL 760  
QY 691 VTAKOLADYLN-----KVNETADALSFPKVGONGDSNNAI-----TVGKDTN 733  
DB 761 AAATSIQGFETNINLVDSHITLVSDDDNIGSGMV-NIDSSSELLFGSTFGILHATLACTG 819  
QY 734 GKTFNTLKLKGNGVNIITNRTAT-----GTVTFGIDQSNGLTTPKLTVGSDTN- 781  
DB 820 SAIVN-----NSANVSLEQASMEFAGTWQVNOGGALT--ASNSNOLGSAKIGLDGTLN 870  
QY 782 GNRVLVIEQVPSADGNTKNIKGLSPTLPSIASPSGRN-----IALGNTIEEKDKSNAASI 837  
DB 871 DNIALENHV--LTNGTILNVAKNLATAFDFGSGFAGFGLVNLTKTTFALSADNAAAL 928  
QY 838 DDVNLNAGFNKNGKDKDFVSTVDVD-FIDG-NATTATVYDEA---NOTSKV--AYDV 890  
DB 929 -----ASATLK-----LSDDSVTTVTGTDRTLHGLDLSGCTLIFDGAVPQSO 944  
QY 891 NVDEKTIETLTDGNGKQKOLGVKTIKLTETSTNGNATTF--TDDHA-----LYKASDIAGN 944  
DB 981 ALNSGTVNIITGSG-----SWNDTDLATNVSLIEQDRAGSTLELINATNVGD 1028  
QY 945 LNTLAEIEHTTKGTANT-----ALQTFVTKVKVDENDK----- 976  
DB 1029 IDALDLVNGTALTSTGTQGVSAIOGGSTVANAIHNYGLASSNSNGDSGLYVNYTSLAL 1088  
QY 977 ---ADDTNATVKGKG--TSQKVNLTCLKKNGLDLTKD-----DGTVT--- 1016  
DB 1089 ELLADGADALLATATESGLTANRVNLNBLFGVGLVWDVAQNCALTLANGSNRYEGTTTVA 1148  
QY 1017 -----FGINTQSLKAGDSTTLNNGLSIKNTASNEQIQVGADGVKFA--MVNNG 1064  
DB 1149 GELILGANGAFQTSLLDIAAGSASANINGVSTQVATNVGTVTGLSGGVLTSGLLTNGG 1208  
QY 1065 VV--GAGID---GTPTRITRDEITGFTNGSLDKSPHLSKDGINAGGKLIINIOSGEI 1117  
DB 1209 ILDLTGALNLTXGASTVAG--GLTGA-GTLNGLNGLSVSAANSGLSGQTHI--ADV 1262  
QY 1118 AKNSHDATVGGKIYDLKTELENKISSTAKTAQNSLHEFVSVADEQGNFTNPNYSYDTS 1177  
DB 1263 A-----SVT-----LTDGTGTSAVEVLGTLNLNCANAAMTVLSGDGTI 1303  
QY 1178 KTSDVITFAGENGITTKVNGVVRVIGIDQTKGLTTPKLTGVNNGKG-----IVINSQ 1232  
DB 1304 NTNAAVTLSGNSFS-----GAHQIGTD-----GELTVQASNLGASSATVNLGTLTS 1351  
QY 1233 QNTITGLSNTLANVTNDKSVRTTEQGNIIKDEKTRAASIVDVLS-----AG 1280  
DB 1352 HLILANGVSESIANVL--GVAGSTV--DIIGADTALTANNSGFLGOYALAGNSKLTVAS 1407  
QY 1281 FNLQNGEAVDFVSYDVTYVNEAN-----GNTTT-----AKVTYDDTSKTSKVYD 1325  
DB 1408 TNNLGAASSVALAGAGDITLSLGFNGTGFNSVTSGLVQVTDADAEVLTSSNGVSNAL-- 1465  
QY 1326 VNVDDTTIEVKOKKLVKVTTLTS--TGTGANKFALSNOAT---GDALVKA--SDIVAH 1377  
DB 1466 -----TIDIADATLNDLIDIALFNHVLTCGLLNVAKNDASTAFDFGTVGAFSGIVNL 1519  
QY 1378 LNT-----LSGDIQTAKGASQANNSAGYVDAGNKVYDSTONKYYQ 1419  
DB 1520 TMTFALSADNAAALARATLKLSDSDSVTTVGAT--DRTLHGLDLNGGTLIFDGSPP--Q 1574  
QY 1420 AKNDGTVDKTKEVAKOKLVA-----QAQTPDGTLAQMNKSVINKNOVDANKKQ 1469  
DB 1575 SQANGVVTVTDALNSGCTISITGAGWNEHHPVTPPNVLSLEQDRCDLILLELINAAVVG 1634  
QY 1470 GINEDNAFVKGLEKASDNKNTKNAAVTVGD--LNAAVQTPLTFTAGDTGTAKKLGTTLI 1527  
DB 1635 NANNLDDLVDGTALTSTGTQGVESAIOGGSTVANAIHNYGLTSSNGSGGLYVNYTSL- 1693  
QY 1528 KGGQTDNKLTDN--NIGVVAGTGTGTVKLAKDLITNLNSVNAAGTKIDKSGISFVDANGQ 1585





QY 932 DHALVKSADIAGNLNLAEEIHTTKGTANTALQOTFTVKKVVDENDKADDTNATVKGDTGS 991  
 Db 1108 DOSLSAK--IKNL-----IGTVAGGA-----TYTLDAINDN---FDLN-----TS 1144  
 QY 992 GRV-----NTLKGKNGLDIKTKDGVTFEIGI---NTQSG-----LKAGUSTLLNNGLS 1039  
 Db 1145 GMVFKHDSILEKNSNTN---DRTITLTSALDPGNNOFIIKLIITDNTKLTIDNNG-- 1199  
 QY 1040 IKNTA-----SNEQIQVGADGVKFEAMVNGV---VGAGIDGTTTRITRD--EIGFTGTNGS 1089  
 Db 1200 --NVATICTANHLK-----QUTFASIDNGALAKVGINVENVTINIKIDIELNEVNANVL 1253  
 QY 1090 LOKSKPHLSKDGINA-----GKKKTIINTQ-----SGETAKNSHDAVTC-----GKIYD 1132  
 Db 1254 FNKNTYTTATNGINGHVDPQGNAGVINLMDIDIEIDGSTVTGDN--VNGTLNPNFGSGKVTG 1311  
 QY 1133 LKTELENKISSTAKTAQNSLHEFSVADEG---NNFT--VSNPYSSYDTSKT----- 1179  
 Db 1312 LINNVIMQAGADVSLASGNSYFTEIOCGNNGNLTFAANSHLTTDINKTGQDNLNVF 1371  
 QY 1180 -----SDVITFAG-----ENGITTKVNGVVRVVID--QT 1207  
 Db 1372 INGSVSGSIGANAAGVDDIIINAGSVNFSNTLKSNGNIVISDGMATMOVNNVTATDISGN 1431  
 QY 1208 KGLTTPKL-----TVGNNGKGIIVINSONGONTIT----- 1237  
 Db 1432 ANNGTLKLNHHTPINTSTLGNNAIG--TIEVANDVTIIGTLQAOIHFSNATQAATLT 1490  
 QY 1238 -GLSNTLVANTDKSGVRTTE-----QGN--IIBDKTKTRAASIVDVLSAGNLOGNG-- 1287  
 Db 1491 LGAASQVNTTITAGNHTHTLEVTFDFTDNGDIIIGDAN--NRLKSI-----ELTNGCTIV 1541  
 QY 1288 --EAVDFVSYDVTNPFANGN-----TTAKVTYDTSKTSKVYDVVNDVTI--EVKDKKL 1340  
 Db 1542 TINSPPHYVSSITTTANNAQGNVKNIEGGITYDLSKIKSLANVQISEDTITIGDVIYSKYL 1601  
 QY 1341 GVKTTTSTGTGTGANKFALSQATGDAV-----KASDIAVHLNLTSGD--IOTAKGASQ 1393  
 Db 1602 NIDAGKTINFDGNNNNPRNLDIPDALIDLVLPRLSLFNFYFTDIKADNLNFADDTAT 1661  
 QY 1394 AN-NSAGVVDAD--CNKVIYDSTDNKY--YQAKNDGTVDKTEKVAKDKLVAQAQPDGTILA 1449  
 Db 1662 ANFKDAVVIDAHIDNGGLFLFNDAWLTQETKNANII-----ELASDKF-----LL 1708  
 QY 1450 QMNVKS-----VINKEOVNDANKKQGINEDNAFVKGLEKASDNKTKNAAVTVG 1498  
 Db 1709 QKNIKAAATLIADNANVLVDNVEVNTNLNVRDVLDA---NYELKVTGNVTHNGLITI- 1764  
 QY 1499 DLNVAQAQPLTFAGDTGTAKKLGTTLTKGGQTDTHKLTDDNNIGVVGAGTGDFTVKLAKD 1558  
 Db 1765 -----ITY---FDTALQGGHILVSQSGSNVDMSDL--DNLI-----IKIAHSD 1803  
 QY 1559 LTNLNS-----VNAAGTKIDEKIGISFVDANGAKANTPVLSANGL----- 1598  
 Db 1804 ITNITSYDKHQHVLKLETAIYTPPTKVIIIDASEEQNKVKWADANGLVLLTDTGGRD 1863  
 QY 1599 DLGGKVIS-----NVGKGTDT--DAANV---QOLNEVRNLLGLGNADGNQVNIADIKK-- 1649  
 Db 1864 DTGGRRDTRGRGNTDNGCRDNGDVGNNSSNEA-----GGSSSDKN--YGITDVPVPIF 1916  
 QY 1650 DPN-----SGSSS---NRTVIAKGTVLGGKGNNDTEKLTATGGVGVGVKDKG-- 1692  
 Db 1917 DSPILDYTKNNYVAGIANQLINHVKGNTTDTAGKLLNDLGFSPNPRVETETDLRLSNR 1976  
 QY 1693 -NANG-----DLNSNVWK-----TQKD--GSKKALLA---TYNAAQGTNYVT 1728  
 Db 1977 INVNGLNEGVSGLNGTEVENFLTDIAINMDFNTAKEIGNRLEELSDANTVNLGNKNTFTLL 2036  
 QY 1729 NNPAAIDRINEGIRFFHVNDGNPEVVGGRNGIDSSAGKSHSVAIGFOAKADGEAAVA 1788  
 Db 2037 NNK-----INLKLNTNNQAIIAAGDE--DNIVTGLWGSFYFKIKQNSKNS-A 2082

QY 1789 IGRQTOAGNQSIAGDINAQATGDSQSIATGCTGNVAG-----KHSIGAIGDSTPVKADNSVSG 1845  
 Db 2083 SGVQSTNGGIIIGFDYNI-----DNSIVIGAAYTMADSKVKHKNDKNGDRTKAKNSIYSI 2138  
 QY 1846 -----NNNQFTDA-----TQDVFVGNNITVTESNVALGNSAIS 1882  
 Db 2139 GLYNWLTNNFFVEAIGVVGGRNKKIKNYEKRTITITDQIAIGKFINTFYSELGCGNYLIS 2198  
 QY 1883 AGTHACT-----OAKKSDGTAGTTTTAGATGTVKGFAG--QTAVGAVSVGASGAER 1931  
 Db 2199 ---HRTIIPMEGRMYATFKNGYKNNNTTFQNLSTIKKNYDKFETILGLNSV----- 2248  
 QY 1932 RIONVAAGEVATSDAVNGSQL-----YKATQSIANATNELDRIHONENKANAGISSA 1986  
 Db 2249 -----THYLSQDIIIIKPELHWFINTYCKNKUPNIDARLD-----GIDEP 2287  
 QY 1987 MAMASMPQAVIPGRSMVTGGIATHN 2011  
 Db 2288 LTTIRFKPAKITYN--LGGGISTKN 2310

RESULT 10  
 Q9VTK8  
 ID Q9VTK8 PRELIMINARY; PRT; 2586 AA.  
 AC Q9VTK8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE CG18331 PROTEIN.  
 GN CG18331.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RA MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.E., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe C.R., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland I.J., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kravitz S., Kulp D., Lai Z.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Liang Y., Lin X.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pohlard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,



Thu Sep 13 14:19:33 2001

```

143 GDVLAEGDASTAIGSD--LYLPKNLDLKNFHEHLKINGHEILKKTOTSTGDKIKYRTR 199
172 G-----ATAGITIOSATPALINLAGAIDGDNNDH-----GREITVTRTSFTGIGCTKL 221
200 AOGHASTAVGAMSYAOGHFSNA-FCTVATAEAYSLAVGLAAQATKOSSIA-VGSNAKAN 257
222 ---HAVT-----FNNGNAGGVSTAGIASATTMTTIGEDAGNVASVVQGLITGAV 267
258 AFAATAGTNTVNLGRGVAGLFGSOILDRDNNTDASAVPLGKTLADQYKATQGGSTD 317
268 NFAAD--GALTANN-----GIAGAVTTANNNTG-----TL-----TVGAGDVTG 304
318 IFSIGNNNNNNSIIRKKIINVAGSRDT-DAVVAQLKLVEELAN-----RKITFGK 368
305 AICV-----NGNILLKQVLFNGASNVATIDATNVT---INNAANVTAAAGAITAAYNFAA 356
369 DGNNSNSVERGLNLTITKGAOTNALTANIGV-----VTGNGLVKVLAKELTG 420
357 DGALTANN---GIAGAVTTANN-NTGTLT---VGAGDVTGAIVGNGNILLKQVLFNGASN 409
421 LTSVATNKITVSTNTNNNAELOSGLTFSPITGTYKTKTVYSIDG-----LKE 469
410 VATIDATN-VTI---NNAANVTAAAGAITAAYN-----FAADGALTANNAGIAGAVTT 457
470 TNDNSIATGTTTITKTKTGAGTNDGVDSEKPYLDNEKLVGNSTLNSGLTNNNTG 529
458 ANNTGTTLTVGAGDVT-----GAIVGNG-NILKQVLFNGASNV--ATIDATVTINNVA 510
530 NKQTOVGA-----NGIKFA-TVANNVANTSATVGTARTTEKIGPAGTN 572
511 N-VTAAGAITAAYNFAADGALTANNAGIAGAVTTANNNTG-TLVGAGDVT---GAIVN 564
573 DGVDQAPYLDKRLKVRVEITDTSGINAGNHKITGLTNGIANTDAVTIKLQKDAKPTL 632
565 GG-----NILKQVLFNGASNVAT---IDATNVTINNNAANVTAAAGITAAYNFAADGAL 615
633 NAGDGI-----SINSNNGDLVDSGNITPTPTNISVTKTLKNSNGTSGNN--KFSVSNH 685
616 TANGIIVCAVTTANNNTGTLTVGAGDVT-----GAIVGNGNILLKQVLFNGA 662
686 DNNSLVTAOLADYLNKV--NETADSLP--SPKVGONGNSNNAITVGKDTNGKTFNTL 740
663 SNVATIDATVT--INNVAANVTAAAGITAAYNFAADGALTANNIV-----707
741 KIKGNGVNIITNTRATGVTFCIDQSNGLTTPKLTIVGSDTNGNRLVIEQVPSADCNSTKN 800
708 -----GAVTTANNNTGTLTVGAGDVTG-----AIG--VNG-----GNILKQ 741
801 II-KGLSPTLPSIASPSGRNIALGNTIEEKDSNAA-----SIDDDLNLNAGNLKNGKD 853
742 VLFNGAS-----NVATIDATNVTINNNAANVTAAAGITAAYNFAADGALTA-----NNG-- 790
854 KDFVSTYDVTDFIDGNATVATVYDEANQTSKVAYDVNVNDEKTLTCDGDKKKQL----- 908
791 -----IAGAVTTAN-----NNTGTLTVGAGDVTGAIVGNGNILLKQVLFNGA 832
909 -GVKTKLTETSTNGNATFTFDDDDHALVKASDIAGNLTALAEETHTTK-----GTANTAL 963
833 SNVATIDATNVTINNNAAN-----VTAAGITAAYNFAADGALTANNIGVAVTTA-- 883
964 OTFTVKVDENDKADDTNATTVGKDGTSKGVNTLKLKNGKGLDIKTDKDGTVTFPGINTQS 1023
884 -----NNTGTLTVGAGDVTGAI-----GVNGGNILLKQ-----VLF--NGAS 918
1024 GLKAGDSTTLNNGLSIKNTAS---NEQIQVGDGVKFAWNVGVVVG-----AGIDGTTRI 1076
919 NVATIDATVTINNVAANVTAAAGITAAYNFAADGALTA--NNGIITGAVTATAGNGT--L 974
1077 TRDEIGFTGNGSLDKSPHLSKDGINAGGKLTINIQSGEIAKNSHDVATGKGIYDLKTE 1136
975 TTAGVGVTVGAVGT-----NAASLKVLN---AKVDNAGAADLVFTSDIY----- 1013
1137 LENKISSTAKTAQNSLHEFSVADEQGNNTVSNPYSDYTSKTSVDITTF-AGENGITTKV 1195

```

---

```

1845 IPNSGRKLNSSSETSTTV---TSSSSSKPQTKYSWSSSSSKKNNGGKNKYWKRWTKK 1901
1710 SKKALLATYNAAGQNTVYNNPAPADIRINEQGRFPHVVDGNQOE-----PVV 1757
1902 SRK-----NNGSSSTIVGE---ESSDSLTDAGVDVDTQGRGLNDEGNSSOSTVTSLLPVV 1952
1758 QGRNIDSASG-----KHSVAIGFOAKADGEAAVAIGROTPAQNOISIAIGDNAQAT 1809
1953 DTSADVQNSSESLTSTENTWTKYSKSKFVKPSKQSSISASKTKTKVTSTSSPPNVKSS 2012
1810 GQOSIAIGTGNVAVKSHGAIGDPTSVKADNSYSVGNNOPTDA-TOTDFGVGNNITVT 1868
2013 SKK-----TSN--SGK---SVKTSSTTITTTSSDPGSSSITGIPONDIKSL-NQVTTT 2061
1869 ENSVALGNSAISAGTHAGTQAKSDGTAGTUTTAGATGVKGF--AGQFAGVAVSVGA 1926
2062 TSSVSQGVPPSSPVVVKTKETSVSKDGKTKRSSTTTTTTTITTKGNSQSGTLTLPADV-GL 2120
1927 SGAERRIQNVAA-----EVSATSTDAVNGSQLYKATQSTANATNELDRIHQNE 1976
2121 KSKTKTTTSTGKTKLSLDLSLPEVDASI--AVNGDESRASIKDNTILSKIDLSLPKLD 2178
1977 N--KANAGISSAMAMAPQAIYPERSM-----VTGGIATHNGOGAVAVGLSK---LSD 2025
2179 ASLVNNGGKSSKSSSTTTTSTTKGNKVSLSLPEVDASIAV-NGDDARSASIKDNTILSK 2237
2026 NGQWVKINGSADTQ 2041
2238 IDLSLPKLDASLVNG 2253

```

---

```

RESULT 11
O9XC47 PRELIMINARY; PRT; 2106 AA.
ID O9XC47
AC O9XC47;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE OUTER MEMBRANE PROTEIN A.
GN OMPA.
OS Rickettsia australis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=787;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PHS;
RX MEDLINE=20487299; PubMed=11034486;
RA Stenos J., Walker D.H.;
RT "The rickettsial outer-membrane protein A and B genes of Rickettsia
RT australis, the most divergent rickettsia of the spotted fever group.";
RL Int. J. Syst. Evol. Microbiol. 50:1775-1779(2000).
DR EMBL; AF149108; AAD39531.2;
SQ SEQUENCE 2106 AA; 209075 MW; A12B28F54BEEDECC CRC64;

```

---

```

Query Match 5.4%; Score 557; DB 2; Length 2106;
Best Local Similarity 21.9%; Pred. No. 2e-15;
Matches 509; Conservative 247; Mismatches 790; Indels 774; Gaps 125;

```

---

```

QY 1 MNHIYKVFINKATGTFMAVAEYAKS-----HSTGGGSCATQGVGSVRTLSFARIAAL-- 52
DB 1 MANISPKLFQKATHKSLKALETTTTPAAILMSSSGAWGVAAGVI-SVNDAAFSNRAVANN 59
QY 53 -AVLIVGATLNGS-----AY-----AQOITKIEIGOT-----NKINNT 85
DB 60 WNETAGGAANGHADGPQDNEAFTYCGNHITTADEAGRIITAINVAGTTPVALNSTONT 119
QY 86 LKGDALATGE---ASTAFGSLSAQGSQAATAGSVKPPDPNNGSNGNNGVSHAGNESIATG 142
DB 120 SVGSIVTGGNLLPVTIADGKSLTGTGKAVA-----ADHGFDAADNNT-GLGRIALG 171

```

```
Db 1014 -----ARTV-----NF-----DIAGGAGVTQVGGNLIATNV 1042
Qy 1196 NKGVVRVGDQTKGLTTPKLTGNNNGKGVINS-----QNGSNTITGLSNTL 1243
Db 1043 NFG-----GNAQGGTLXLPVGVKSYTLSTGLTIANGNATINI-NTL 1082
Qy 1244 ANVTNDKSVRTEOGNIIKDEKTRAASI-----VDVLSAGFNLOGNEAVDFVSTYTV 1299
Db 1083 GLVTABNAXIGTVAQINI--QDNKTFAINKNADIEIL-----NAQIDFGKANSKL 1132
Qy 1300 NFANGNTTAKVTYDDTSTKSVVYDVNVDDTTIEVKDKKLGKVTITLTSTGTGANKFAL 1359
Db 1133 FLVNSAT-----DDRVTIKN-----DLPFATGGMMLL 1163
Qy 1360 SQAATGDALVKASDIVAHLNLTSGDIOTAKGASQANNSAGYVDADGN---KVIYDSTDNK 1416
Db 1164 FG-----TANLITLQGD-----GAXKALGTAGNKLASLNVLGKVAENNID-- 1203
Qy 1417 YQAKNDGTVDKTEKAVDKLVQAQATPD-GTLAOMNVKSVINKQOVNDANKKQGINEDN 1475
Db 1204 -----TTNLVAFNIINTTFQVDVGGIT--NOISVIN-----IGAANASLGL---- 1242
Qy 1476 APVKGLEKASDNKTKNAAVTVGDLNVAQATPLTFAGDTGTTAKKLGELTLTIKGGQTDN 1535
Db 1243 -----PAAAGSYTIDAN-----GGWVGILAN--GQTINF--AHEDAE 1275
Qy 1536 KLTDNIGVAGTGTGTVKLAKDLTLN-----SVNAG-----GTKIDBK 1575
Db 1276 LVLQNS--AAGNSTITLNAALDPCAPNKGKLVADSGXAGGRVILASVRNATYGTAVNKL 1332
Qy 1576 GLSFVDANGQAKNPVLSANGLDILG-----GKVIS--NVGKGTDKTDAAVVOOLNE 1625
Db 1333 KLEFRGNTQFQIDIF-ANNLEFPVPAATYNNKINDXNLSFGAATLTQNGNI----- 1386
Qy 1626 VRNLLGLGNDNAD-GNOVNIADIKKDPNSSSNTVVKAGTVLGGKGNNDTEKATGGV 1684
Db 1387 -----NGNVDFNQAAVITL-----GANKNIT--GSVTSNGVNGT-TIATGAS 1427
Qy 1685 QVGVDKDNANGDLNVMVVK-----TQDKGSKALL--ATYNAAGOT 1724
Db 1428 TI-----NGHI?NIAMLKVAGAVSITQGGNTSITEIQNGTALLTLPANFNLTG-- 1477
Qy 1725 NVVTNPAEADRIEIQGRFHVNDGNOEPVVOGRNGIDSSASCKHSAVIGFAKADGE 1784
Db 1478 -----SINKTGOALKLNFTNGGSRGVV-----GT 1503
Qy 1785 AAVAIGROTAQAGNSTAIGDNAQAT-----GDQSI A--IGTGNVVAGHSGAIGDPSTVK 1837
Db 1504 AANSVGDIITTTGTTNFPASVNAKGAATLCCTTSPADTPTNTGAVTLAKAS----- 1553
Qy 1838 ADNSYSVGNNOFTDATQTDVFGVGNNTIVTESNVAL-----GNSAISAG-----THAG 1888
Db 1554 -----ITNFAKNVATSF--AANNATINFGNSLAENSNITGSGTTITLCSAQVITYG 1603
Qy 1889 T-----OAKSDG-----TAGTTTACATGT-----VGFAGQATV 1919
Db 1604 TGSFTDTLTNTFPDGAASKDGHILKSGSTLDSGVSTLALVVTATNPDINNIPNTRY 1663
Qy 1920 GAVSGASCAERRIOWAAGEVSATSD-----AVNGSOLYKATOSTANATNELDHR- 1971
Db 1664 TVISAETAG--LKPTPAGNVKVTNNDRNRVNTFDESTLTILFAEDIAEEVIEDFRP 1720
Qy 1972 -----IHQNEKANAGISSAMAMSPQAYIPGRSMV-----TGGIATHINGQAV 2016
Db 1721 GGPLANIPNAANV--KKSLELMQAPIGSDACLAFDNMGKLNLPQOADAVENTHMQDVV 1776
Qy 2017 AVGLSKLSDNGOWVK-----INGSADTQGHVGAAGVAG 2050
Db 1777 KPSDTIAAINNOVILSNISSSSLINLARMADAIQPAAVAG 1816
```

RESULT 12

```
Q9JY30
ID Q9JY30 PRELIMINARY; PRT: 2514 AA.
AC Q9JY30;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE HEMAGGLUTININ/HEMOLYSIN-RELATED PROTEIN.
GN NMB1768.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Utterback J.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RA "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT WC58."; 287:1809-1815(2000).
RL Science
DR EMBL; AE002526; AAF42109.1;
DR TIGR; NMB1768;
SQ SEQUENCE 2514 AA; 265615 MW; 95643A671B3BC268 CRC64;
```

Query Match 5.3%; Score 553; DB 2; Length 2514;

Best Local Similarity 20.0%; Pred. No. 3.8e-15;

Matches 503; Conservative 332; Mismatches 95; Indels 782; Gaps 121;

```
Qy 4 IYKVFENKATGTFMAVEYAKSHSTGGSCATQGVGS--VRTLSFARIAALAVLVICATL 61
Db 5 LYRVIPNRKRGAVVAETTKRE---GKSCADSDSGSAHVKSVPFG----- 47
Qy 62 NGSAVAQIQTGKTIEQTNKINNLTKGDA--LATGEASIAFG-----SLSKAQGSQAI 112
Db 48 -----TTHAPVCRSNIFSFSLGFSICLAVGTANTAFADGIADRAKAPKQATIL 98
Qy 113 ATGSKVPDPN-----NGSNGNVGSHAKGNESIAIGGDV-----LAEG 149
Db 99 QTCNGIPQVNIQTPPTSAGVSVNQYAOFDVGNRGAILNNSRSNTQTQLGGVIQGNPWLAR 158
Qy 150 DASIAIGSDDLLYLPKNLDLKNPEHLKHGHEILKKIQTSTDGKIKYRTPAOGHASTAVG 209
Db 159 EARVVV-----NQINSSHSSQMGNYIEVGGRAEVVIANPAG 195
Qy 210 AMSYAQGHFSNAPGTATAEAYSLAVGLAAQATKQSSIAV---GSNAKANAF----- 259
Db 196 IAVNGGFFINASKATLTTCQPOQYQ-AGDLGFKPIRGVNVVIAHGHLDARDTDFRILSYH 254
Qy 260 -----AATAIGGNTVNVNLGRGVALGFGSQILDRDNT-DASAYVPL-----GK- 301
Db 255 SKIDAPWQDVRVAGQNDVWATGNAHSPILNNAANTSNNTANNNGTHIPLAIDTGLK 314
Qy 302 -----TLADQYKATROGDSITDIFSIGNSNNNNSIRKLIINVG-----AGSRDTD 346
Db 315 GGMVANKITLISAEQAGIRNQG---OLFA--SSGNVAIDANGRLVNSGTPMAAANKD 369
Qy 347 AVNVAQLKLVEELANKITFKGDGDNNSNSVERGLNLTIKGDAQTNALTEANIGVWTD 406
Db 370 -----NTAEHKVNIRSGVENS----- 387
Qy 407 GNLKVKVLAKELTGLFSVATNKITVSTNTNNNAELQSGLTFSPTGTKDTKTVSIDG 466
Db 388 -----TAVSQQGTQIHQSQTQNTGTLSSGEILLHNSGLKNETSGTIEA 432
Qy 467 LKFTDNSNATKG---TTRITKKKIGFACTND-----GVDESKPYLDNEKLVGKSTLN 518
```

Thu Sep 13 14:19:33 2001

Db 433 ARLAIDTDLNNOGKLSQTSOKLHIDAQKMDNRGRMLQDTAPTASNGS---SNQTCN 489  
 Qy 519 SGLTVNNTTGNKQIOVGANGKFAFVANNVANTATVCTARITEKIGFAGTNGDVQDE 578  
 Db 490 SYNASFHSSTTTPTTATG-TGTATVSNISNITPTFAD-GTIR-----THGALDMS 537  
 Qy 579 APYLDKERLAVGVEIITDSGI-NAGNHKITGLT-----NGTANTDAVYIKOLKDA 628  
 Db 538 GSITAN-----GQTVSAQOGLNAGGIDHQLNKAAGSAFNDHNGTIIISDAVHIQ----- 587  
 Qy 629 KPTLNAGDGISINNSNODLVDSGNITPTYNISVTKL----- 668  
 Db 588 -----AG-SLNQON-----GNITT-RQLEIETDQDNHAGKLLSABEADLAIVS 631  
 Qy 669 --NSNGTSGNNKFSVNAHD-----NNSLVATKDLA-----DYLKNVNETADSAL 711  
 Db 632 LNNQGEIATNQOLI--IHGQOOSTAVIDNTNGTIQSGRDVAIOAKSLSNNGTLLAADNKL 689  
 Qy 712 -----PSFKVQNGDSNNATVKGDTNGKFTNKLKGENGVNITNRACTVTFGIDQS 766  
 Db 690 DIALQDDFVVERNIVAGNELSL--STRGSLKNSHTLOAGKRIRKAKNLDNAAGNI-QS 746  
 Qy 767 NGLTTPKLTIVGSDTN-GNRLVIEOVPSADGNSTK-----NIIKGLSPTLPSTIASPSGR 818  
 Db 747 GGT-----DIGTQHNLNTRGLI-----DGOQTKIAGOMNIGTG-----RIYGD 787  
 Qy 819 NIALGNT-IEEKDKSN---AASIDVLNAGNLNKKDKDFVSTYDTVDFIDGNATTAT 874  
 Db 788 NIAIAATRLDNQDENGTAATAARENLNLGIGOLNRENLSIYS-----GNDMAGV 838  
 Qy 875 VTYDEANQTSKAVDYNDVEKFIETLGDNGKKGKLVKFI-----KLTETSTNGNAT 925  
 Db 839 GALTNGQATGAQRIHNAAGTIEAA---GKMLGVEKHLNTHNEHLKQLVETGRE-HIV 894  
 Qy 926 TETSTDDHALVKAS---DIAGNL-NLAEIHTTKGTANTALQFTVVKVDENDKADDT- 980  
 Db 895 DYEAFGRHELLREGTQHELGSWVDESJHLTPDCAAHENHKKYDYBKVTQKVTQTA 954  
 Qy 981 -----NAITV-GRD--GTSKQVNTLKGKGLDIDTKDG---TVTFG---INTQSG 1024  
 Db 955 PAKIISGNDLTIDGKRVFNTDSOI-----IAGGN-LIVQTEKDLGNEQTFGEKKVSENG 1009  
 Qy 1025 L-----KAGDST-----TL-----N 1034  
 Db 1010 KLHSYWRKHKGSDTSGHSEQNTPLPEITRNISLGSFAYESHKALSHIAPSQOTELPQ 1069  
 Qy 1035 NGLSKNTASNEQIOVGADGVKAFMNVNNGVAGIDGTTRIT--RDEIGFTGTNGSLDK 1092  
 Db 1070 SNGISLPTNSFTPLPSSSLYIINPVNKGYL---VETDPFANYROWLGSYMLDSLKL 1126  
 Qy 1093 SKPHLSK---DGINAGGKKTINISQGEIAKNSHDVATGKKIYDLKTELENKISSAKTAK 1149  
 Db 1127 DPNLHRLRGDGYE--QRLNEQIAEL--TGHRRLDG---YONDEEQFKALMDNGATAA 1179  
 Qy 1150 NSLHFSVADEQGNFTVNPVSSYDTSK--TSDVITFAGENGITTKVNGVVRVGVDOFK 1208  
 Db 1180 RSM-----NLSVGIALSAEQVAQTSIDVILVQKE-----VKLPDGGTQ 1218  
 Qy 1209 GLTTPKLTIV-----GNNGKGIIVNSQONTITG-----LSNTLA 1244  
 Db 1219 TVLVPOVYVVRVKNGDIDGKCALLSGSTQINVSGLSKNSGTIAGRNALINTDTLONIGG 1278  
 Qy 1245 NVTNDKGSVTRTEQ-----GNIIKEDKTRAASIVDVLISAGFNLCGNEAEVFSVYDVTN 1300  
 Db 1279 RIHAQSAVATQDINNIGGLMSAEQTL-----LNAGNNI----- 1314  
 Qy 1301 FANGNTTAKVTYDDTSTKSVWYDVNVDDTTIEVKDKILGVKVTTLTSTGTGANKFALS 1360  
 Db 1315 --NSOSTTASSONTGSGST-----YLRMAGIYI-----TGKEKGVLA 1350  
 Qy 1361 NOATGDALVKASDIVAHNLNTLSGDIQAKGASQANNNG-----YVDAD 1404  
 Db 1351 AQAGKD-----INIITAGISNSEQOTRLQGRDINLDTVQTSKHQATHEDAD 1399

Qy 1405 GNKVIYDSTD--NKYQAKNDGTVDKTEVAKDKLVAQAQTPDGTILA-----OMNVKSVI 1457  
 Db 1400 -NHVIRGSTNEVGSIQTKGDTVILLSGNLL--NAKAAEVSSANGTLAVSAKNDINISAGI 1456  
 Qy 1458 NKEOVNDANKKQGINEDNAFVKLEKAAONKYNKAAVTVG-----DLANAQTP 1508  
 Db 1457 NTHVDDASKHTRSGGONKLIVITDKAQSHHETAQSSTFECKQVVOAGNDANILGNSVI 1516  
 Qy 1509 TFAGD-----TGTTAKK-----LGETLTIKGGQTDNKL 1537  
 Db 1517 SDNCTQIOAGNHVRIGTQTOSQSEYHQTKSGLMSAGIGFTIGTSKNTQENQSQSNEH 1576  
 Qy 1538 TDNNIGVAVAGTGTVKLAKDLTNLSVNA---GGTKIDEGISFVDANGAKANTP--- 1591  
 Db 1577 TGTSTVGLKGDIT--TIVAGKHVEIGSTVSPPEGNNTIYAQSIDIAAHNKLNSNTQTY 1634  
 Qy 1592 -----VLSANGLDGLGVIS-----NVGKGJKD-----TDAANV--QQLNEVRLNLLGL 1632  
 Db 1635 EOKGLTVAFSPVTDLAQAATAVAOSSKQVGSKNDRYNAMAANAAGHQAQVOTCKSAQNL 1694  
 Qy 1633 GNDNADGNQVNIADIKDPNSGSSNR--TVIKAGTV-----LGGKNNDEKLTATGGV 1684  
 Db 1695 ANGTNNAKQVSI-----ITYGEQONROTTVQANQAQASQIQAGKK-----TTLIATGAA 1745  
 Qy 1685 Q-----VGVDKDGNA-----NGDLSNVVKTOKDGSKKA-----LLATYNA 1720  
 Db 1746 EQSNINIAGSDVAGKAGTILADIADNDITLQSAEQSN---TERGQNKSGAGNAGAVSEGO 1801  
 Qy 1721 AGQTNVVT--NPAEALIDRINEQIRFFHVNDGNOBPVQVGRNGIDSSASGKHSVAIGFO 1778  
 Db 1802 GWSLGVTAGGNVKGKYGNGDSITHRSHIGDKGSQTLIQ--SGGDTTIKGAQVRGKGVQ 1859  
 Qy 1779 AKA-----DGEAAVAIGROTQAG---NOSIAGDNAQATGDSIAI 1816  
 Db 1860 VNAKLSIOSVDRETYQSKQNASAQVTVGYGFSAGGYSQSKIRADHVSVTEQSGIYA 1919  
 Qy 1817 GTG--NVVACKHSGATGD--PSTVKADNSVSGVNNQFDTATQ---DVFQVGNNTVTVTES 1870  
 Db 1920 GEDGYQKVNHNHDLKGGIITSTQSAEDK---GKNRFOTATLTHSDI---KNHSQYKG 1971  
 Qy 1871 NSVALGNSAISAGT--HAGTQAKKSDGTAGTTTACATGTVKGFAGTAVGAVSVCSAG- 1928  
 Db 1972 ESFGLGASASISGKTLGGGAQNKPNKHILTSAADKNSASSVGT--GSDSDSQSSTIKSGI 2030  
 Qy 1929 -----AERRIQNVAAGEVSATSDAVNGSOLYKATQSIANNNTN-----ELD-- 1969  
 Db 2031 NTRNIQITDEAAQIRLTGKTAQTKADIDTNTVTDTAERHSGSLKNTFNKEAVQSELDLQ 2090  
 Qy 1970 -----HRIHONENKANAGISSAM-----AMASMPQAVIPGRSMVTGCIATHNGQAV 2016  
 Db 2091 RTVSQDFSKNVQQAQTEINQHLDKLKADKAEATAAALANG--DMETAKRKAHEADAA 2149  
 Qy 2017 AVGLSKLSDN---GOWVFKINGS---ADTQHVGAAYGA-----GPHF 2053  
 Db 2150 A-----KADNNQOQKVILNMLASGLAAPTQSGAGIAAATASPASVAYIGQHF 2196

RESULT 13

Q3ZHL0  
 ID Q9ZHL0 PRELIMINARY; PRT: 4919 AA.  
 AC Q9ZHL0:  
 DT 01-MAY-1999 (TReMBLrel. 10, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)  
 DE LARGE SUPERNATANT PROTEIN 2.  
 GN LSPA2.  
 OS Haemophilus ducreyi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_Taxid=730;  
 RN [1]  
 RP SEQUENCE FROM N.A.







```
Db 834 TNPVVVTGALDNTGN-----ANNQIVTFTGDSVTGNGINGNATLATSIVA 879
Qy 1205 DQ--TKG-----LTPPKLT-----VGN--NNGKGVINSQNGQNTIT 1237
Db 880 GRATLGGAIIKATTTKLTYNASAVTFNNVVVGTGAIDNTGNANG--IVTFTGDSVTI 935
Qy 1238 G-LSNVLANTNDKSGVRTEQNIKIDKEDKTRAASIVD-----VLSAGFNLOGNEAVD 1291
Db 936 GNIGNTALATVNVGAVTQLQAGSLD-----ANNIDFGARSTLEFNGPLDGGKAIP 988
Qy 1292 FVSTYDTVFANFANGNTTAKVTYDDTCK-----TSKVYVDVNV-----DDTTIEV 1335
Db 989 Y----YFKGAINGANNAILNVNFKSLTASHLTGTVAEINIGACNLFAIDASAGDVILNA 1045
Qy 1336 KD---KKLGKVTTLTSTGTGANKFALSNOATGDALVKASDIVA-----HLNLTSGDIQT 1387
Db 1046 QDIHFRALDSALVSLNLTGCVNNILL-----AADLVPAGVDEGTVVFDGCV-- 1092
Qy 1388 AKCASQANNSAGV-----DADGNK-----VIYDS---TDNKYQ----- 1419
Db 1093 -NGLTSGNVAGARNIGDVGNGKNTLLIYNAVTTTDDVNLGIONVLINNADFTSST 1151
Qy 1420 AKNDGTV---DKTKEVAKDLVAAQOTPDCTL--AQMNVKSVINKEQVNDANKKQGINED 1474
Db 1152 AFNAGHIQINDATYTI--DANNGLMIPAGNIQFAHADAOILILNSSGNDRTITLGANID 1209
Qy 1475 NAFVKGLEKAASDNKTNAAVTVGDLNVAQAOTPLTFAGDTGTTAKKLGLTLTIKGGQT-- 1532
Db 1210 -----PDNDEGIVI-----LNSVT-----AGKLTIAAGCKTEG 1238
Qy 1533 DTNKLTD-----NNTGVVAGTDGFTVKLAKDLTLNLSNVNAGGTKIDKGISFVDANGQA 1586
Db 1239 GAHKLODIVFKGSGDFG--TAGTFTNTNTIVLDITGOLGATTANV-----VLEKDA---- 1289
Qy 1587 KANTPVLNGLDGLGKVISNVGKGTDKTDAANVQOLNEVRNLLGLNDNADCNQVNIAD 1646
Db 1290 ----VOLTGT--NIGGFDFENAKNGT-----VTLN-----NNNVAG 1321
Qy 1647 IKKDPNNGSSNRVIAKGTVLGGKGNNDTEKLA---TGGVQVGVKDKGNAN--GDLSNVW 1702
Db 1322 TVK--NTGNTNGTLLI---VLCASNLNRVNGIAMLKVGAGNVYIAKGGNVKIGEIQGTG 1375
Qy 1703 VKTKDCKSKALLATYNAAGOTNYVNTNPAEADIRNEOGIRFFHVNDGNQBPVVOGRNG 1762
Db 1376 TNT-----LTLPAHPKLTG-----SINKTGGQALKLFNMGSGVSGW----- 1413
Qy 1763 IDSSASGKHSVAIGFOAKDGEAAVIGROTAQGNOSIATGDNAAQT---GQDSIA--- 1815
Db 1414 -----GTAAHSVGDITTAGATSPASSVNAKGTATLGGTTSFANTF 1453
Qy 1816 IGTGNVYVAGKHSGAIGDPSTVKADNSYSVGNNOFTDATQDTDFGVGNNTVTTESNSVAL 1875
Db 1454 TNGAVTLAK-----GSIISFAKNVTATSF--VANSATINFNGSLAF 1493
Qy 1876 GSNALSATGTHACTQAKKSDCTAGCTTTTACA-----TGTVK-----GFAQOTAVGA 1921
Db 1494 NSNITGS-----GTTTLGANQVTVYTGSGFTDTLTTLNTTFDGAARSGG 1537
Qy 1922 VSYGASABRRIQNVA--AGEVSATSTDAVNGSOLYKAT 1958
Db 1538 NTLKSGSLDLSGVSNLALVVTATNFDMMNISPDTKYAT 1576
```

## RESULT 15

```
P94750
ID P94750 PRELIMINARY; PRT; 2349 AA.
AC P94750;
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-MAY-2001 (TReMBLrel. 16, Last annotation update)
DE INVASIN (FRAGMENT).
OS Escherichia coli.
```

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.

OX NCBI\_TaxID=562;

RP [1]

RC SEQUENCE FROM N.A.

RC STRAIN-K12;

RX MEDLINE-97251358; PubMed-9097040;

RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,

RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,

RA Mizobuchi K., Mori H., Motomura K., Nakade S., Nakamura Y.,

RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,

RA Sivasubaram S., Tagami H., Takeda J., Takemoto K., Wada C.,

RA Yamamoto Y., Horiuchi T.,

RA "A 400-Kb DNA sequence of the Escherichia coli K-12 genome

RT corresponding to the 40.1-50.0 min region on the linkage map.";

RL DNA Res. 3:379-392(1996).

DR EMBL; D90836; BAA15799.1; -.

DR InterPro; IPR001639; -.

DR InterPro; IPR003344; -.

DR InterPro; IPR003535; -.

DR Pfam; PF02369; B19\_1; 13.

DR PRINTS; PR00810; BCTERIALGSPC.

DR PRINTS; PR01369; INTIMIN.

FT NON\_TER 2349 2349

SQ SEQUENCE 2349 AA; 247433 MW; E45F5FEE2AE76B59 CRC64;

## Query Match

5.1%; Score 526.5; DB 2; Length 2349;

Best Local Similarity 22.0%; Pred. No. 4.5e-14;

Matches 523; Conservative 302; Mismatches 933; Indels 615; Gaps 125;

Qy 47 ARTAALAVLVIGATLNGSAYAAQITTK--TEIGQTNKIN-----NTLKGDALATG 94

Db 68 AQIAIANANTVPVTLGSAESQAQSVARFGISVAELRKLNQFRTPARGFDNVQGDLDV- 126

Qy 95 EASIAFGSLSKAQS-----QAIA-----IGSVKPPNNGSNCNGVSHAKGNESIAIG 143

Db 127 PAQVSEKKLTPPPGNSSDNLEQOIASTSQOIGSLAEDMNSE--QAANMARGWASSOAS 184

Qy 144 ---DVLAE--GDASTAIGSDDLPLPKNLDKNEPHKLHG-----HEILKK 184

Db 185 AMTDWLSRFCTARTILGVDE----DFSLANSQDFLHPWYETPDNLFQSOHLHRTDER 239

Qy 185 IQSTSD-GTKTKYRTRAQG-----HASTAVGAMSYAQGHF---SNAFGTYATA 228

Db 240 TQINNGLGWRHFTPTWMSGINFFFDHLSRYHSRAGIGA-EYWRDYLKLSNGYLRLTNW 298

Qy 229 EAAVSLAVGLAAQNTKOSSIAVGSNAKANAF--AATAIGGNTVYN--LGRGVALGFGSQI 284

Db 299 RSAPELDNDYEARP-----ANGMDVRAESWLPAMPHLGKLVYEQYYGDEVAL-FDKD- 350

Qy 285 LDRDNTDASA---YVPLG-KTILADQYKATROGDSDFISI-----GNS-----NNNN 328

Db 351 -DQSNPHALTAGLNYTPPLMTFSABQROKQOGENUTREAVDFTWQPGSAMQKLOPNE 409

Qy 329 SSTIRKTIIVGASRDPTDAV-----NVAQLKVLBELANRKITFKGDDGNNNS-NSVERGLGN 383

Db 410 VAARRSL-----AGSR-YDLVDRNNNIVLEVRKKELVRLTLTDPVTGKSGEKSLSVSLQT 464

Qy 384 TLTIKG-DAQTNALTEANIGVVTGNGLKVKL-AKELTGLTSVSAT--NKITVSTNTNNN 439

Db 465 KYALKGYNVEATALEAAGGKVVTGDKDILVTLTPAYRFTSTPTDNTWPIEVTAEDVAGNL 524

Qy 440 AELQSGGLTSPITGKTQVYSIDGLKFTNDSNSIATKGTTRITKKKGK-----FA 492

Db 525 SNREQSMVVVQAPTLQSKDSSV-SLSTQTLNADSHSTAT--LTFIAHDAAGNVPVGLVLS 581

Qy 493 GTNDGVDE--SKPYLDNEKLVGNSTLNSGSS--LTVNNTTGNKQIQVGANGI---KFA 543

Db 582 TRHEGVQDITLSDWKN-----GDGYSYQILTTGAMSGTLTLPOLNGVDAAKAP 631

Qy 544 TVANNAV-NTSATVGTARITEEK-----IGFAGTND-GVDEQAPYLDK-----ERL 587



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 23:04:48 ; Search time 33595.7 Seconds  
(without alignments)  
3196.157 Million cell updates/sec

Title: US-09-361-619-10

Perfect score: 6942

Sequence: 1 atgaatcacatctataaagt.....ttggtgcaggttttcacttt 6942

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_ba3: \*  
4: gb\_in1: \*  
5: gb\_in2: \*  
6: gb\_in3: \*  
7: gb\_om: \*  
8: gb\_ov: \*  
9: gb\_pat1: \*  
10: gb\_pat2: \*  
11: gb\_ph: \*  
12: gb\_pl1: \*  
13: gb\_pl2: \*  
14: gb\_pl3: \*  
15: gb\_pl4: \*  
16: em\_ba1: \*  
17: em\_ba2: \*  
18: em\_fun: \*  
19: em\_htgo\_hum: \*  
20: em\_htgo\_inv: \*  
21: em\_htgo\_rod: \*  
22: em\_htg\_hum1: \*  
23: em\_htg\_hum2: \*  
24: em\_htg\_hum3: \*  
25: em\_htg\_hum4: \*  
26: em\_htg\_hum5: \*  
27: em\_htg\_hum6: \*  
28: em\_htg\_hum7: \*  
29: em\_htg\_hum8: \*  
30: em\_htg\_inv1: \*  
31: em\_htg\_inv2: \*  
32: em\_htg\_other: \*  
33: em\_htg\_rod: \*  
34: em\_hum1: \*  
35: em\_hum2: \*  
36: em\_hum3: \*  
37: em\_hum4: \*  
38: em\_hum5: \*  
39: em\_hum6: \*  
40: em\_hum7: \*  
41: em\_in: \*  
42: em\_om: \*  
43: em\_or: \*

44: em\_ov: \*  
45: em\_pat: \*  
46: em\_ph: \*  
47: em\_pl: \*  
48: em\_ro: \*  
49: em\_sts: \*  
50: em\_sy: \*  
51: em\_un: \*  
52: em\_vi: \*  
53: gb\_sts1: \*  
54: gb\_sts2: \*  
55: gb\_sts3: \*  
56: gb\_sy: \*  
57: gb\_un: \*  
58: gb\_vil: \*  
59: gb\_viz: \*  
60: gb\_htg1: \*  
61: gb\_htg2: \*  
62: gb\_htg3: \*  
63: gb\_htg4: \*  
64: gb\_htg5: \*  
65: gb\_htg6: \*  
66: gb\_htg7: \*  
67: gb\_htg8: \*  
68: gb\_htg9: \*  
69: gb\_htg10: \*  
70: gb\_htg11: \*  
71: gb\_htg12: \*  
72: gb\_htg13: \*  
73: gb\_htg14: \*  
74: gb\_htg15: \*  
75: gb\_htg16: \*  
76: gb\_htg17: \*  
77: gb\_htg18: \*  
78: gb\_htg19: \*  
79: gb\_htg20: \*  
80: gb\_htg21: \*  
81: gb\_htg22: \*  
82: gb\_htg23: \*  
83: gb\_htg24: \*  
84: gb\_htg25: \*  
85: gb\_pr1: \*  
86: gb\_pr2: \*  
87: gb\_pr3: \*  
88: gb\_pr4: \*  
89: gb\_pr5: \*  
90: gb\_pr6: \*  
91: gb\_pr7: \*  
92: gb\_pr8: \*  
93: gb\_pr9: \*  
94: gb\_rol: \*  
95: gb\_rod: \*  
96: gb\_in4: \*  
97: gb\_pr10: \*  
98: em\_ba3: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6942	100.0	6942	10	AX079922 Sequence
2	2939.2	42.3	6159	10	AX079920 Sequence
3	2852.4	41.1	62909	10	AX067457 Sequence
4	2590	37.3	6141	10	AX079918 Sequence
5	2590	37.3	6972	10	AX079917 Sequence
6	2580.6	37.2	6973	9	AR040716 Sequence
7	2580.6	37.2	6973	10	AX079913 Sequence
8	2485.8	35.8	5976	10	AX079914 Sequence

9	2484.8	35.8	6259	10	AX079924	Sequence
10	1752.4	25.2	3135	10	AX079957	Sequence
11	1421.8	20.5	1448	10	AX079959	Sequence
12	90.8	1.3	3900	3	PMU277635	
13	90.8	1.3	10533	1	AE006194	Pasteurel
14	82.8	1.2	3294	10	I55122	Sequence 1
15	82.8	1.2	3700	3	HU38617	Haemophilus
16	81	1.2	11159	1	AE005587	Escherich
17	81	1.2	28090	2	AP002566	Escherich
18	78.4	1.1	149789	60	AC007926	Trypanoso
19	76.6	1.1	11650	3	U32846	Haemophilus
20	76.4	1.1	7062	3	U41852	Haemophilus
21	72.8	1.0	3324	3	STY277623	Salmonell
22	70.2	1.0	15021	1	AE003982	Xylella f
23	69.8	1.0	5738	10	I55123	sequence 3
24	68	1.0	14235	59	HZU67264	Helicoverpa
25	68	1.0	13403	58	AF271059	Helicove
26	64.4	0.9	8139	3	PMU277636	Pasteurel
27	64.4	0.9	10903	1	AE006108	Pasteurel
28	63.8	0.9	10259	1	AE004017	Xylella f
29	61.4	0.9	144506	77	AC084438	Homo sapi
30	61.4	0.9	160896	88	AC025540	Homo sapi
31	61	0.9	7408	6	DVLP612P	Z29565 D.virilis L
32	60.2	0.9	199551	60	AC006281	Plasmodiu
33	59.6	0.8	15861	4	AE001413	Plasmodiu
34	58.4	0.8	60	9	AR040717	Sequence
35	58.4	0.8	1797	2	AF226381	Neisseria
36	57.8	0.8	171798	4	AC084397	Trypanoso
37	57.6	0.8	8504	5	AF030693	Plasmodiu
38	57.6	0.8	11682	1	AE005230	Escherich
39	57.6	0.8	262278	2	AP002552	Escherich
40	57	0.8	11534	96	PFAARPLPR	Y08926 P.falciparu
41	57	0.8	164399	96	PFMAL13P6	Z98551 Plasmodium
42	57	0.8	318221	84	PFMAL13P3	AL049184 Plasmodiu
43	56.8	0.8	1770	2	AF226362	Neisseria
44	56.8	0.8	1770	2	AF226363	Neisseria
45	56.8	0.8	1770	2	AF226384	Neisseria

ALIGNMENTS

RESULT	1					
LOCUS	AX079922	6942 bp	DNA	PAT	22-FEB-2001	
DEFINITION	Sequence 10 from Patent WO0107619.					
ACCESSION	AX079922					
VERSION	AX079922.1	GI:13159441				
KEYWORDS						
SOURCE	Moraxella catarrhalis.					
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;					
REFERENCE	1 (bases 1 to 6942)					
AUTHORS	Loomsore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.					
TITLE	Recombinant high molecular weight major outer membrane protein of moraxella					
JOURNAL	Patent: WO 0107619-A 10 01-FEB-2001;					
FEATURES	CONNAUGHT LABORATORIES LIMITED (CA)					
source	Location/Qualifiers					
BASE COUNT	2286 a 1621 c 1564 g 1471 t					
ORIGIN						

Query Match 100.0%; Score 6942; DB 10; Length 6942;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6942; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 atgaatcacatctataaagtcatctttaacaagccagcacatttatggcgtggca 60  
|||||

1	ATGAATCACATCTATAAAGTCATCTTTAAACAAAGCCACAGGCACATTTATGGCCGTGGCA	60
61	gagtgcccaatccacagggaggagtagaccgagcaggtgggcagc	120
61	GAGTGGCCAAATCCACAGCGGAGGAGTAGCAGTAGTACCGCAGGACAGGTGGCAGC	120
121	tctctgtcatcgctgactgctgtgtgccagctcgctatacttagtgatcggtgcgacg	180
121	TCTCTGTCTATCCGCTGACTGCTGTGGCAGCTCGTATCTAGTAGTCTGGTGCAGC	180
181	ctcaatggcagtgcttatgctcaataataacagacgcatctgggtaccacaggaac	240
181	CTCAATGGCAGTGCTTATGCTCAAAATAATAGCAAGATCCCATTTGGTACCAACAGCAAC	240
241	aatgacaatgctcggtcagcagcaatgaagcatcattgctggttagtgctgtaagca	300
241	AATGACAAATGCTCGGTGAGCAATGAAGCATCCATTTGTTGGTAGTCTTGTAAAGCA	300
301	catgccaatcaagctattgctatcggtggttagcaaacagatcctcgtaacagcgct	360
301	CATGCCAATCAAGCTATTGCTATCGTGTGTAGCAAAACCAAGATCCTCGTANTCAGCGGT	360
361	aatcaaaagcaggttcccaacgcaaaaggtaaagagtcctatcgccatcggtggtgatga	420
361	AATCAGAAAGCAGGTTCCTCAGCGCAAAAGGTAAAGAGTCCATTCGCCATCGGTGTGATGA	420
421	ctggtgaggtgatgctcctcgatgctcctggttagtgagtagtatttggataggaat	480
421	CTGGCTGAGGGTGATGCTCCTGATTCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	480
481	agcctaactctaaatcccaatggctctcttagcactcttattcaaaacacacagta	540
481	AGCACTAATCTAATATCCAAATGGTCTTCTTAGCACTCTTATTTAAAACCATACAGTA	540
541	ttaagcacaatcacgagactcaaatggcttcagaaatataagacacagcagcagagga	600
541	TTAGCCCAATACGAGACTCAATGGTTCAGAAATATAGACCCACAGCAGCAGCAAGGA	600
601	cagccagctactcagtgaggagccatgcatatgcaaaagggtcatctttgccacgctctt	660
601	CAGCCAGTACTGCTGAGGAGCCATGGCATATGCAAAAGGTCTATTTGCCAGCGCTTT	660
661	ggtacacggtcagcagctgaaggcaactattccttggcagtaggtcttaccgcaagcc	720
661	GGTACACGGTCAACAGCTGAAGGCAACTATTCTTGGCAGTAGGTCTTACGCCCAAGCC	720
721	gaaaagagatatacaatcgctattggttcaatgcaagctatcaattatggagcacta	780
721	GAAAAGGATATACAATCGCTATTGCTTCTAATGCACAAGCTATCAATATATCGACACTA	780
781	gcccctgggtgcagatcagctgagtgattgattacggtattgcccctagggttatggtct	840
781	GCCCTTGGTGCAGATACGAGTTGATTGGATTACGGTATTGGCTTATGGCCCTAGGTATG	840
841	cagatccttaataataataataataataataataataataataataataataataata	900
841	CAGATCTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	900
901	tcaaacataaaatcgtctaaagccacccggcgaatggtttatttccattggtagtagcact	960
901	TCAACATAAAATCGTCTAAAGCCACCGCAATGGTATTATTTTCCATTGGTAGTAGCACT	960
961	atcaagcgtaaaatcaatgctcggtcaggttatgagagatccgagtcggtgcaatg	1020
961	ATCAAGCGTAAAATCAATGCTGGAGATCTGGTATAGCGTCAAAATTACTTTTAAAGGTG	1020
1021	gcacagctaaaagcgtggagaaatcgtggaagcgtcaaatctacttttaaggtgatgat	1080
1021	GCACAGCTAAAAGCGGTGGAGATCTGGTATAGCGTCAAAATTACTTTTAAAGGTGATGAT	1080
1081	aacggttactgggttaagaaaaaacctggcgagactttaaccattaaaggttgtagacc	1140
1081	AACGGTACTGGCGTTAAGAAAAAACCTGGCGCAGACTTTAAACCATTTAAAGGTGAGACC	1140







```
Qy 5521 atcaagcgaataaataacagcagcatcacccctaagcaaaatgggctgaatgtaggcgtt 5580
Db 5521 ATCAAGGCAAGATAACAGCAGCATCACCCCTAAGCAAAAGATGGGCTGAATGTAGGCGGT 5580
Qy 5581 aaggtcatagcaaatgggttaagcacaataaagacacacagcactgccaatgtacaacag 5640
Db 5581 AAGGTCAATACAGATGGGTAAAGGCACAAAGACACACCGCCTGCCAATGTACAACAG 5640
Qy 5641 ttaaacgaatacgaacttgttgggtcttggtaactgtgttaataacgcgtgacgccc 5700
Db 5641 TTAACCAAGTACGCAACTTGTGGGTCTTGGTAATGCTGTGAATGATAACGCTGACGGC 5700
Qy 5701 aatcaggtaaacattgocgcacatacaaaaagacocaaattcaggtttcateatcacaacgc 5760
Db 5701 AATCAGGTAACATATGCGGCACATCAAAAAGAGCCAAATTCAGGTTTCATCATCTAACCGC 5760
Qy 5761 actgtcatcaaaagcagcagctactgtggttaaggttaataacgataccgaaaaactt 5820
Db 5761 ACTGTCAATCAAGCAGCAGCTACTTGGCGGTAAAGGTAAATACGATACCGAAAACTTT 5820
Qy 5821 gccactgtgtgtgtacaaagtggcggtgataaaagacggaacgcataacgcgtatttaagc 5880
Db 5821 GCCACTGTGTGTGTACAAAGTGGCGGTGATAAAGACGCGCAACGCTAACGCGGATTTAAGC 5880
Qy 5881 aatgtttgggtcaaaaccccaaaagatggcagcaaaaaagccctgtcgcacactttaac 5940
Db 5881 AATGTTTGGGTCAAAACCCCAAAAGATGGCAGCAAAAAGCCCTGTCTGCCACTTTAATAC 5940
Qy 5941 gccgcaggtcagaccaaactatttgacaacaacacccgcagaaagcattgcagaataaat 6000
Db 5941 CCCGCAGGTCAAGCAAACTATTGTACCAACAAACCCCGCAGAAAGCCATTGTACAGAAAT 6000
Qy 6001 gaacaaggtatccgctcttccatgtcaacgattggaatcaagagcctgtggtacaagg 6060
Db 6001 GAACAAGGTATCCGCTCTTCCATGTCAACGATGGCAATCAAGAGCCTGTGTACAGGG 6060
Qy 6061 cgtaacggcattgactcaagtgcctcaggaagcaactcagtgcggaatgtttccaggcc 6120
Db 6061 CGTAACGGCATTTGACTCAAGTGCTCAGGCAAGCACTTCAGTGCGGATAGGTTTCCAGGCC 6120
Qy 6121 aaggcagtggtgaagccgctgtgcatagcagcagcaaaaccccaagcagcacaacatcc 6180
Db 6121 AAGGCAGATGGTGAAGCCGCGCTTGCCATAGGCAGACAAACCCCAAGCAGCAACCAATCC 6180
Qy 6181 atcgccatcgtgtataacgcacagccacgagcgatcaatccatcgccatcgtgtacagcc 6240
Db 6181 ATCGCCATCGGTGATACGACACAAAGCCACGCGGATCAATCCATCGCCATCGGTACAGGC 6240
Qy 6241 aatgtgttaacaggttaagcactgtgtgcatcggcagcccaagcactgtttaaggctgat 6300
Db 6241 AATGTGTAAACAGGTAAGCACTCTGTGTGCCATCGCGACCAAGCACTGTTAAGGCTGAT 6300
Qy 6301 aacagttacagtggtgtaatacaaacacagtttatcgatgccactcagacagcagtgcttt 6360
Db 6301 AACAGTTACAGTGTGGGTATATAACAACAGTTTATCGATGCCACTCAACACGATGCTTTT 6360
Qy 6361 ggtgtgggcaataacatcacccgtgacccgaagtaactcgtgtgcttaggttcaaaactct 6420
Db 6361 GGTGTGGCAATAACATCACCGTGAACCGAAAGTAACTTCGCTTTAGGTTCAAACTCT 6420
Qy 6421 gccatcagtgaggaacacacgcagggcacaagaagcgaataatctgcagcagcaggg 6480
Db 6421 GCCATCAGTGAGGACACACGAGGACACAAAGCCAAAATCTGACGGCACAGCAGGT 6480
Qy 6481 acaaccacacagcagtgcaacaggttacgtttaagcgttttctggaacaagcgggtt 6540
Db 6481 ACAACCCACACAGAGTGCAACAGGTACCGTTAAAGCGCTTTGCTGACAAACCGCGGTT 6540
Qy 6541 ggtgcggtctccgtgggtgctcaggttgctgaacgcgcgtatccaaaatgtggcagcaggt 6600
Db 6541 GGTGCGGTCTCCGTGGGTGGCTCAGGTGCTGAGCGCGGTATCCAAAATGTGGCAGCAGGT 6600
```

```
Qy 6601 gaggtcagtgccaccagcaccgcatcggttcaatgttagccaggtgtacaagccacccaa 6660
Db 6601 GAGGTCAGTGCCACAGCAGCATCGGTCAATGTTAGCCAGTTGTACAAAGCCACCCAA 6660
Qy 6661 ggcattgccaaacacaaatgaatgagcttgaccatcgatccaccaaaacgaataaaagcc 6720
Db 6661 GGCAATGCCAAACCAACCAATGAGCTTGACCATCGTATCCACCACCAAAATAAAGCC 6720
Qy 6721 aatgcagggatttcacagcagatgagcgatggcgtccatgccacaagcctacatcctggc 6780
Db 6721 AATGCAGGGATTTTCATCAGCGATGGCGATGGCGTCCATGCCACAAGCCTACATCTCTGGC 6780
Qy 6781 agatccatggttaccgggggtattgcccacacacacacacacacacacacacacacacac 6840
Db 6781 AGATCCATGTTTACCGGGGTATTGCCACCCACACACACACACACACACACACACACAC 6840
Qy 6841 ctgcgaagctgclcgataatggtcaatgggtatttataatcaatggttcagccgatacc 6900
Db 6841 CTGTCGAAGCTGTCGATAATGTCATGCGTATTTAAATFCAATGTTTCAGCGGATACC 6900
Qy 6901 caaggccatgagggcgccaggttggtgaggttttcacttt 6942
Db 6901 CAAGCCATGAGGGCGCGCATGTTGTCAGGTTTTCACTTT 6942

RESULT 2
AX079920
LOCUS AX079920 6159 bp DNA PAT 22-FEB-2001
DEFINITION Sequence 8 from Patent WO0107619.
ACCESSION AX079920
VERSION AX079920.1 GI:13159440
KEYWORDS
SOURCE
ORGANISM Moraxella catarrhalis.
Bacteria: Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 6159)
AUTHORS Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.
TITLE Recombinant high molecular weight major outer membrane protein of moraxella
JOURNAL Patent: WO 0107619-A 8 01-FEB-2001;
CONNAUGHT LABORATORIES LIMITED (CA)
FEATURES
source
1. 6159
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 2035 a 1386 c 1385 g 1353 t
ORIGIN
```

Query Match 42.3%; Score 2939.2; DB 10; Length 6159;  
Best Local Similarity 70.2%; Pred. No. 0;  
Matches 5085; Conservative 0; Mismatches 768; Indels 1395; Gaps 29;

```
Qy 1 atgaatcacatctataaagtcactcttttaacaaagccacagcagcatttatggcgtggca 60
Db 1 ATGAATCACATCTATAAAGTCATCTTTAAACAAAGCCACAGSCACATTATGGCGGTGCG 60
Qy 61 gagtgcgcaaaatccacagcagggagtagcagtagtaccgcagcagcaggtggcgagc 120
Db 61 GAATATGCCAAATCCCAACAGTACCGGGGGGTAGCTGTCTACAGGGCAAGTTGGCAGT 120
Qy 121 tctctgtcatccgcctgactcgtgttgccacgctcgtatctcctagtgcgtgcgacg 180
Db 121 GTAGGCACCTTAAGCTTTGCCCGTATTGCCGCGCTCGTCTGCTCGTATGCGTGCAGCG 180
Qy 181 ctcaatggcagtgcttatgctcctatgcctcctcctcctcctcctcctcctcctcctcct 237
Db 181 CTCAATGGCAGTGCTTATGCTCAACAAATTAATACCAAGATCGAAATTTGTCACAAACAA 240
Qy 238 aa-----caatgacaatgcctcggctgagcaatgaagcatccattgctatt 282
Db 241 AAGATAAACACACGCTGAAAGCGGATGCCCTAGCGAGGTGAAGCATCAATGCTTTT 300
```



Db 2413 TTGTCCTCCAACTGCTAGCATTTGCGAGTCCAAAGTGGC---CGCAACATAGCACTGGGC 2469  
QY 2236 aatacaatcaccagcgacgaagacaaaatccaaagcccgagatcgcgtgatattataat 2295  
Db 2470 AATACAAAT---CGAAGAAAGACAAATCCACGCTGCCAGCATGTGATGTGCTAAAT 2526  
QY 2296 acaggctttaacctaaaaataatagcaactccggttggtcttctccactataaact 2355  
Db 2527 GCAGGCTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 2586  
QY 2356 gtgactttatcagtgcaatgccaccacgcctaagtaacttacgatgaacaaacaa 2415  
Db 2587 GTTGACATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2646  
QY 2416 accagtaagttaacttatgatgtcaatgtgtagagaaacacattgaactcacagcgat 2475  
Db 2647 ACCAGTAAGTGGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2706  
QY 2476 aatggcaagacaaacaaattggcgtaacaaacacacactgacacaaacaaatgcta 2535  
Db 2707 AATGGCAAG---AAACAACCTTGGCGTCAAAACCATCAAACTGACCGAAACAAAGTACTAAT 2763  
QY 2536 ggtaaagcaacaaactttagtaaccacgcataaagatgccttggtaacgcacaaagacatc 2595  
Db 2764 GGTAAATGCAACTACATTTAGTACCGAGATGACCATGCCCTTGTAAAGCCAGTGATATC 2823  
QY 2596 gccgaataatcacaaccccttagcgaagaaattcacaccacaaagacagacagcacc 2655  
Db 2824 GCCGGCAATCTAACACCCCTAGCGAGGAAATTCACACCACCAAGGACAGCAAAACACC 2883  
QY 2656 gccctacaaacctttaagtcacaaagacggtgcaactgatga-----cgaa 2703  
Db 2884 GCCCTACAAACCTTTTACCGTTTAAAGGTAGATGAAATGAAAGGTGATGACACCAAC 2943  
QY 2704 accatcccggtgggtaagatgttacacaaagcgaagaccgtcaacactctaaactc 2763  
Db 2944 GCCATACCGTGGGTAAAGATGGCACA-----AGTGGTAAAGTCAACACCTTAAACTC 2997  
QY 2764 aaaggtgaaacgcttaacggttgctaccacataaagatggtacggttaccttggcatt 2823  
Db 2998 AAAGGTAAAAAGCTCTTGATATTAAAAACCGACAAAGATGCTACGGTTTACCTTTGGCATT 3057  
QY 2824 aacaccaaagcggtcttaaaagccgagcgaagacacacactctaaacaaagatggtctct 2883  
Db 3058 AACACCCAAAGCGGCTTAAAGCCGGCGACAGCACCACTCTTAAACAAATTTGGTCTCT 3117  
QY 2884 attaaaccccgctagtaacgaacaaatcgaagtcggtgctgtagtgcgtgaatttgc 2943  
Db 3118 ATTAAAAACCGCTAGTACGAACAATCCAACTCGGTGCTGATGGCGTGAAGTTTGGCC 3177  
QY 2944 aaggttgataaggtgtaattcaagcactggcattgtaggcacaaagccgtatcacaaagat 3003  
Db 3178 ATGGTTAATAATGTTGTTGATGGTGGTGGCATTGATGGCACAACTCGCATTTACCAGAGAT 3237  
QY 3004 caaattggcttactggggctaaatggctcacttgatataccacacaaacccacactaacaaa 3063  
Db 3238 GAAATGGCTTTTACTGGGCAATTAATGGCTCACTTGTATAAAAAGCAACCCCACTAAGCAA 3297  
QY 3064 gacaagcttaagtggtgaagtigaaattaccaacactggcattaacgcaggttgtaaa 3123  
Db 3298 GAC-----GGCATTAACGCAAGGTGGTAAA 3321  
QY 3124 aagattaccacattcaatcaggtgatattaccacaaacagcaatgatgctgtgacagggc 3183  
Db 3322 AAGATTACCAACATTTCAATCAGGTGAGATTGCCAAAAACAGCCATGATGCTGTGACAGGC 3381  
QY 3184 ggtcgggtttatgatttaaaaaacgcgaacttgaaagcaaaatcaacagtgctgtgtaaaaca 3243  
Db 3382 GCGAAGATTTATGATTTTAAAAACCGCACTTTGAAATAAAAATACAGTACTTGCACAAACA 3441  
QY 3244 qcacaaacattacacgaattctcagtagcagataacaaaggttaacatcactttacggtt 3303

Db 3442 GCACAAAACCTCATTTACACGAATTTCTAGTAGCAGATGAACAAAGGTAATAACTTTACGGTT 3501  
QY 3304 agtaaccttactccagttatgacacctcaaaagccctctgatgtcatcacactttgcaaggt 3363  
Db 3502 AGTAACCTTTACTCCAGTTATGACACCTCAAAAGACCTCTGTATGTCTATCACCCTTTGCAGGT 3561  
QY 3364 gaaaacggcattaccacaaaggttaataaaaggtgtgtgctgtggcgattgacacaaacc 3423  
Db 3562 GAAAAACGGCATTTACCACCAAGTAAATAAGGTGTGTGCGTGTGGCATTTGACCAACC 3621  
QY 3424 aaggtcttaaccacgcctaagctaacggtgttaataataatggaaggaagcatgtcatt 3483  
Db 3622 AAAGGCTTAAACACCGCTTAAGCTGACCTGGGTAAATAATATGCAAAAGGCATTTGTCATT 3681  
QY 3484 gacagtaagaagtgtcaaaaataccatcacaggaactaaagcaacactctagctaaatgttacc 3543  
Db 3682 AACAGCCAAATTTGTCAAATACCATCACAGACTAAGCAACACTCTAGCTTAATGTTTACC 3741  
QY 3544 aatga-----tggtgcaggacacgcgactaagccaaagggcttgccaatgacacccgacaaa 3597  
Db 3742 AATGATAAAGGTAGCGTACGCCACACACAGAGGCAATATAATCAAAAGACCAAGACAAA 3801  
QY 3598 accggtgccgcagcattggtgatgtgctaaacgcagggctttaacttgcgaaggaatggt 3657  
Db 3802 ACCGTCGCCGCACGATTTGATGTGCTAAGCGCAGGCTTTAACTTGCAGAGCAATGGT 3861  
QY 3658 gaagcggttgaacttctccacttatgacactgttgaactttatcgatggcaatgacacc 3717  
Db 3862 GAACGGTTGACTTGTCTCCACTTATGACACCGCTCAACTTTGGCAATGGCAANTACCACC 3921  
QY 3718 accgtaaggtgacacttatgatgacacaaagcaaacccaggttaagtggtctatgatgtaaat 3777  
Db 3922 ACCGCTAAGGTGACTATGATGACACAAAGCAAAACCAAGTAAAGTGGTGTATGATGTCAT 3981  
QY 3778 gtgataataaaacccattgaagtgcacaaagtataaaacacttgcgtcaaacaccacaca 3837  
Db 3982 GTGGATGATCAACACCTTGAAGTTAAA---GATAAAAAACTTGGCGTAAAAACCCACACA 4038  
QY 3838 ctgacacaaacaaagtcgtaattgtaataatga---accataatttagtgcgcgagtggtg 3894  
Db 4039 TTGACCACTACTGGCACAGGTTGCTAATAATAATTTGCCCTTAAGCAATCAAGCTACTTGGCGAT 4098  
QY 3895 gcccttgttaagccagtgatatacgcaccccatctaaacttggctggcgacatccaa 3954  
Db 4099 CGGCTTGTCAAGGCGAGTATATGCTGCTCATCTATAACACCTTATCTGCGCAGATCCAA 4158  
QY 3955 accgcaaaaggggcaagcgaagcagctcagcaagctatggtatgctgtagtggcaac 4014  
Db 4159 ACTGCCAAAGGGGCAAGCCAAAGCAACACTCAGCAGGCTATGTGGATGCTGTGTCGAAT 4218  
QY 4015 aaggtcactatgacagtaccgataaagaagtaactatcaagtaactgacaaaggggtcaagt 4074  
Db 4219 AAGGTCACTATGACAGTACCAGTAACCAAGTACTTATCAAGCCAAATAATGATGGCAGATT 4278  
QY 4075 gacaaaaacaaagagttgccaagacaaactggtcgcacaaagcccaaacccccagatggc 4134  
Db 4279 GATAAAAACAAAGAGTGGCAAAAGACAAACTGCTGCGCCCAAGCCCAACCCCAAGATGGC 4338  
QY 4135 acattggtcctaaatgaatgtcaaatcagtcattacaaaaggaagtaaatgagtcacaa 4194  
Db 4339 ACATTGGCTCAATGATGTCAATCAGTCAATTAACAAAGAACAAAGTAATATGATGCCAAT 4398  
QY 4195 aaaaagcaagcatcaatgaagacaaacgcctttatcaaaagggcttgaacacgcgcgcaaa 4254  
Db 4399 AAAAAAGCAAGCATCAATGAAGACAAACGCCCTTTGTTAAAGGACTTTGAAAAAGCCGCTTCT 4458  
QY 4255 gacacaaaaacaaaaacgcgcagtaactgtgggtgatattaaatgcggttgcacaaaca 4314  
Db 4459 GATAACAAAAACCAAAACGCCGCGAGTAACCTGTGGGTGATTTTAAATGCGGTTGCCCAACA 4518  
QY 4315 ccgctgaccttgcaggggatacagggcacaacgcgtcaaaacactggcgagactttgacc 4374  
Db 4519 CCGCTGACCTTTGCGAGGGATACAGGCACAACGGCTTAAANCTGGGCGAGACTTTGACC 4578





```
Db 5752 GCGGTTGGTGCCTGCTCCGCTGGGTGCTCAGGTGCTGAACGCCGTATCAAAATGTGGCA 5811
QY 6595 gcaaggtaggtcagtgccaccagcacccgatgcggtcaatggtagccagttgtacaaaagcc 6654
Db 5812 GCAGGTGAGGTGAGTGGCCACCAGCAGCGATGCGGTCAATGGTAGCCAGTTGTACAAAGCC 5871
QY 6655 accaaggcatggcacaacccaatgagcttgacccatgctatccaccacaacgaaat 6714
Db 5872 ACCCAAGCATTTGCCAACCAACCAATGAGTTGACCATGTTATCCACCAACGAAAT 5931
QY 6715 aaagccaatgcagggttccatcagcagtagcgatggctccatgccacaagccatatt 6774
Db 5932 AAAGCAATGCAAGGATTTTCATCAGCGATGGCGATGGCGTCCATGCCCAAGCCTACATT 5991
QY 6775 cctggcagatccatgattaccgggggattgcccaccacaacggtcgaaggtgcgggca 6834
Db 5992 CTGGCGAGATCCATGTTACCGGGGTATTGCCACCACCAACGGTCAAGGTGCGGTGGCA 6051
QY 6835 gtgggactgcaagctgcgataatgcaatggatgcaatgggtatttaaatcaatggttcagcc 6894
Db 6052 GTGGGACTGCGAAGCTGTCGGATATGTCATATGGGTATTTAAATCAATGGTTTCAGCC 6111
QY 6895 gatacccaaggccatgaggggcaggtggtgaggttttcaattt 6942
Db 6112 GATACCAAGGCCATGATGGGCGCGAGTTGGTGCAGGTTTCACPTT 6159

RESULT 3
AX067457 AX067457 62909 bp DNA PAT 24-JAN-2001
DEFINITION Sequence 32 from Patent W00078968.
ACCESSION AX067457
VERSION AX067457.1 GI:12545077
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 62909)
AUTHORS Lagace,R.E., Patterson,C. and Berg,K.L.
TITLE Nucleotide sequences of moraxella catarrhalis genome
JOURNAL Patent: WO 0078968-A 32 28-DEC-2000;
Incyte Genomics, Inc. (US)
FEATURES
Source
1. .62909
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 18449 a 13798 c 12744 g 17918 t
ORIGIN

Query Match 41.1%; Score 2852.4; DB 10; Length 62909;
Best Local Similarity 69.3%; Pred. No. 0;
Matches 4853; Conservative 0; Mismatches 991; Indels 1156; Gaps 25;

QY 1 atgaatcacatctataaagtcatctttaaaagccacagggcacatttatggcctggca 60
Db 42630 ATGAATCACATCTATAAAGTCATCTTTAACAAAGCCACAGGCACATTTATGGCCGTGGCG 42689
QY 61 gagtgcgcaaatcccaagc-ggagggagtagcagtagtacgcagcagcaggtgggca 119
Db 42690 GAATATGCCAAATCCACAGCAGCGGGGGGATAGCTGTGCTACAGGCAAGTTGGCAG 42749
QY 120 ctctcgtcaccgcctgactcgtgttgcacgctcgctatctcctagtcaggtgcag 179
Db 42750 TGTATGCACCTCTGAGTTTTCGCCGCTATTGCCGGCCTCGCTGCTGCTGATCGCTGGCAG 42809
QY 180 gctcaatggcagtgcttatgctcaaaataatagcaagatgcgcatattggtaccacaggcaa 239
Db 42810 GCTCAGTGGCAGTGCTTATGCTCAAGGTA---TCGCAATTTGGTCAACAAACAACTC 42866
QY 240 caatgcaatgctcgtggctagcaatgaagcatccattgctattggttagtcttgaagc 299
```

```
Db 42867 AAAAGCGATGCGAGCGACAGGTAAATCAATCCGTGCTTGGTGAAAAAGCTGTGC 42926
QY 300 acatgccaatcaagctatgctatcggtgtagcaaaaccagatccctcgtaataacagcgc 359
Db 42927 AAGTGCAGCTCAAGCCATCGCCATCGGTGCTAGCCAAACAAATCAATATAA----- 42977
QY 360 taatcagaaggcaggttcccacgcccaaaagtaaaagagtcctccatcggtggtgatgt 419
Db 42978 TAGTAAAGTATCAGGTGCCAACGCCAAAGGTGAAGAGTCCATCGCCATCGGTGATGT 43037
QY 420 actggctgagggtgagctcgtattgccattggttagtgatgacttatatttgataggaa 479
Db 43038 ATTGGCTCAGGGTCCCTGCGATTCGCATCGGTAGTAGTATATATTTGGAGAGTAA 43097
QY 480 tagcactaacctetataatataccaaatggtctcttagcactcttatcaaaaccatacagt 539
Db 43098 AGG-----CTTGTATACACTCCCAAGAGATTAATAAACTTATTGAAAAACATTGAGA 43151
QY 540 attacgccaaatagca-----gactcaaatggtcttcagaaatatagcgcacagcgc 593
Db 43152 ATTAACAGAAATACGAAAACTGGATCCCAATAGTAATCAAAATAATTTACACACAAAGC 43211
QY 594 agaagacacagccagctactgcagtgaggagccatggcatatgcaaaaggtcattttgccaa 653
Db 43212 ACAAGGACACGCCAGTATTTCAGTGGGAACCATGGCAGGAGAGGGTCAATTTGCCAA 43271
QY 654 cgccttggtagacggtcaacagctgaagcgaactatccttgccagtaggtcttaccgc 713
Db 43272 CGGTTTTGGTACACGGGCAATAGCTAAAGCAACTATTCCTTGGCAGTGGGTCTTACCGC 43331
QY 714 caaagccgaaaaggatatacaatcgctatttggttctaataatgcacaagctatcaattgg 773
Db 43332 CAAAGCGGACACAGAAATCTTCAATCGTCTTGGTTCTTAATGCACAAAGTAGCAAGTTGC 43391
QY 774 agcactagcccttggtgcagatactcgagttgattggttagtcaggtatttcctaggtta 833
Db 43392 AGCGATAGCCATTGGTGGAAATTAATACTGAGTTGGGTGCGAGGTGTTCCCTTAGTTA 43451
QY 834 tggttctcagatccttaataataataataataataataaagccttataccagag 893
Db 43452 TAATTCTCAGATC---ACTGCTAATGATAATANTCTAAATGGAGCCATATAAACACAGCG 43508
QY 894 taatgggtcaaacataaaatcgctctaaagccaccggcaat-----gg 935
Db 43509 TGAACATTTACACAGGATCATGAAGCCACAAAAACGGTCAGCGCAAGAAATACAGTGA 43568
QY 936 ttattttccattgtagtagcactatcaagcgtataaataatcaatcaatgctcggtcaggtta 995
Db 43569 TATATTTTCCATTGGTAATAGTGATCTCAGACGTAGAAATCATCCATGTCGGTGCAGGTGT 43628
QY 996 tgaggataccgatgcggtcaatgtggcacagctaaaagcgggtgagaatctggctaagcg 1055
Db 43629 TGATGCCACCGATGCGGTCAATGTGACACAGCTAGAAGCGGTGTTGAACCTGGGCTAAGAA 43688
QY 1056 tcaaatcactttaaagggtgtagtaaacggtactggtggttaagaaaaaactggcgagac 1115
Db 43689 GCGTAAATTTACTTTTAAAGGTTGATGACAGTACTGCGGTACAAAAAGGTTTCGACGAGGA 43748
QY 1116 tttaaccattaaagggtgtagaccacaagcgacaagctaacgataataataacattgg 1175
Db 43749 GTTAACTATTAAAGGGGGGCGCAAAATGCCGGTGAATTAACCGACAAATAACTACGCTGT 43808
QY 1176 tgt---ggtaacagataataactggtctgaaagtctaaacttgctaaaaacccataagcgg 1232
Db 43809 GGTAAAGAGATAATGATAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 43868
QY 1233 tcttgaacaggttagcaccacaaaacccataaccgcagcgagaaaggttacggtta-----gg 1286
Db 43869 TCTTGAAGCAGTAAACACGCAAAACCTTAACCGCACAGGTGCAATCACCGTAAATAATAA 43928
QY 1287 tagtgtaataaacacgctgagctacaaagcgtggttttaaccttttaaccccccaacaaa 1346
```





Qy	3486	cagtaagatggttcaaaataccatcaacaggactaagcaacacactctagctaatgtaccaa	3545
Db	46053	CAGTAAGATGGTCAAAATACCATCAGAGGACTAAGCAACACTCTAGCTATGTTAGCAA	46112
Qy	3546	tgaaggcagacacacactaaagcgaaggcttgcacagacacacacacacccgtgc	3605
Db	46113	TGATGGTGCAGGACAGCAGCTAAGCCAAAGGGCTTGGCGATAACACACCGAACCCCGTGC	46172
Qy	3606	cgcagcatttggtgatgtgtctaaacgcaggctttaaacttgcagaagcaatggtgaagcgg	3665
Db	46173	TGCCAGCATTTGGTGATGTATTAATGCAAGCGTTTAACTGCAAGCAATGGTACAGCGGT	46232
Qy	3666	tgactttgtctccacttatgacactgttgacatttcgatgagcgaatgccaccacgcgtaa	3725
Db	46233	TGACTTTGTCFCCACTTATGACACTGTCAACTTTCGCGATGGCAATGCCACACCGCTAA	46292
Qy	3726	ggtacacttatgacacacagcaaaacagtaaaagtgtctatgatctcaatgtggataa	3785
Db	46293	GGTGACCTTATGACAGCAGCAAAACCAAGTCAAGTGGTCTATGATGTCAATGTGGATGA	46352
Qy	3786	taaaacattgaagtgcacagtgataaaaaaacttggcgtcaaaaaccacacacttgacaa	3845
Db	46353	TACAACCATTTGAAGTTAAA---GATAAAAACTTTGGCTAAAAACCAACCATTTGACCG	46409
Qy	3846	acaagtgctaatggttaatgca---acaaatttagtgcgcgcgatggcgtgccttgt	3902
Db	46410	TACTGGCACAGGTGCTATAAATTTGGCCTAAGCAATCAAGCTACTGCGATGCGCTGT	46469
Qy	3903	taagcgaagtgaatcgccaccatcataaacttaccttgctggcgacatccaaaccgcgaa	3962
Db	46470	CAAGGCCAGTATCGTTGCTATCTAATCAACACCTTATCTGGCGACATCTCAAACTGGCAA	46529
Qy	3963	aggggcaagccaagcaagcagctcagcaagctatgttgatgctgatggcaacaaggtcat	4022
Db	46530	AGGGCAAGCAACGCAACAACTCAGCAGGCTATGTGGATGCTGATGGCAATAAGGTCAT	46589
Qy	4023	ctatgacgtaccgataagaagtagtactcaagtccaatgacaagggctcaagtggacaaaaa	4082
Db	46590	CTATGACAGTACCAGTATACAACTACTACTCAAGCCAAATAATGATGGCAGATGTGATAAAC	46649
Qy	4083	caagaagtgcacaaacaaactggctgcacagcccaacccacagatggccatitggc	4142
Db	46650	CAGAAGTGTCCAAAGCAAACTGGTGCGCCAAGCCCAACCCACAGTGCACTTGGC	46709
Qy	4143	tcaaatgaatgtaaatcagtcattaaacaaagagcaagtaaatgatgccataaaaaagca	4202
Db	46710	TCMAATGAATGTCMAATCAGTCATTACAAAGACACAGTAATGATGCCATAAAGCA	46769
Qy	4203	aggcatcaatgaagacaacgcctttaacaaaggcttgaaaaacccgcacaaagacaccaa	4262
Db	46770	AGGCATCAATGAAGCAACAGCCTTTGTTAAAGGACTTGAANAAGCGCTTCTGTATAACA	46829
Qy	4263	aacacaaacccgcagtaactgttgggtgatttaaatgcccgttgcacaaacacgcgtgac	4322
Db	46830	AACCAAAACGCCGAGTAACCTGTGGTGATTTAAATGCCGTTCGCCAAACACCGCTGAC	46889
Qy	4323	ctttgcaggggaacagggcacaacgcgtcaaaaaactggcgagactttgacatcaaaag	4382
Db	46890	CTTTCGAGGGATACAGGCACACAGCGCTAAAAACTGGCGAGACTTTGACCATCAAAGG	46949
Qy	4383	tgggcaaacagacaccaaataagctaacccaataacacatcggtgtggtagcagtagctga	4442
Db	46950	TGGGCAACAGACACCAATAAGTAAACCGATAATAACATCGTGTGGTAGCAGTACTGA	47009
Qy	4443	tggcttcaactgtcaaaactgtccaaagacctaaacaaatcttaacagcgttaatgcaggtgg	4502
Db	47010	TGGCTTCACTGTCAAACTTGGCAAGACCTTAACCAATCTTAACAGCGTTAATGCAGGTGG	47069
Qy	4503	caccagaattgatgaaaaagcactctctttgttagacgcaaacgcgtcaagccaaagcaaa	4562
Db	47070	CACCAAAATTTGATGCAAAAGCGGTGCTCTTTTGTAGACTCAAGCGGTCAAGCCAAAGCAA	47129
Qy	4563	caccctgtgctaagtgccaatgggctggacactgggtggcaaacgcgactcagataacatcgg	4622

Db	47130		TACCCCTGTGTAAGTGGCAATGGCTGGACTGGGTGGC-----	47169
Qy	4623	tgacgtgttgatgataaacagatgcggtgaactttaagcagtttaatgaagtgcgcaaac	4682	
Db	47170	-----	47169	
Qy	4683	ggtcaacacacccaaacacaaagtaactcagggtgcattaccctttgtgtaacgga	4742	
Db	47170	-----	47169	
Qy	4743	tgcgaatggcgaagcccatcaatgcaccgatggcaagccccaaaaagccatcaaggggcgc	4802	
Db	47170	-----	47169	
Qy	4803	cgatgggtaaatactatcacgcacacgcgcaacgcggtacctgtggacaaagatgcaagcc	4862	
Db	47170	-----	47169	
Qy	4863	catcacgcatggacaaacttgcgaatctgcagctcatggcaaaccccttgatgcagg	4922	
Db	47170	-----	47169	
Qy	4923	tcataacgctgtgcgaagcctagtcggaactcagatgccatcccttaacccaacatcaa	4982	
Db	47170	-----	47169	
Qy	4983	gtccactttgccacaaattgacacaccacacaggttaatgccaatgcaggcgaagccca	5042	
Db	47170	-----	47169	
Qy	5043	aagtctgccagcctatcagcagcacagcaagaagtaatgtgccagtgtcaagaatgtgct	5102	
Db	47170	-----	47169	
Qy	5103	aaatgtaggctttaacttgcagagccaatcacaatcaagtgaacttgcgaagcctatga	5162	
Db	47170	-----	47169	
Qy	5163	taccgtcaactttgtcaattgtacaggtgcgacatcacaaagcgtgcgtagtctgatgg	5222	
Db	47170	-----	47169	
Qy	5223	cacgatgagtaacatcacccgtcaacacgcgccttagcagcgcgatgatgtgccaatgt	5282	
Db	47170	-----	47169	
Qy	5283	gcttatcaagccaaaagatggtgaattctcaaaagcagacgacctcatgcgaacggctc	5342	
Db	47170	-----	47169	
Qy	5343	actaaaagcagcgaataatcagccagtgatgccaaaactccaactggctaaagccttgtaa	5402	
Db	47170	-----	47169	
Qy	5403	ccccaatgctgtaaaaggcagtaacaggcgtatgcagtggctcttaataacttatcaaaagc	5462	
Db	47170	-----	47169	
Qy	5463	ggtatttaaatccaaagatggtacaactactaccacagtaagctctgtggaatcagtat	5522	
Db	47170	-----	47169	
Qy	5523	ccaaggcaaaagataaacacgacgacatcacccctaaagcaagaatgggtggaatgtaggcggttaa	5582	
Db	47170	-----	47171	
Qy	5583	ggatcatcagcaatgtgggtaaaggcacaaaagacaccgacgtgcgaatgtacaacagtt	5642	
Db	47172	GGTCATCAGCAATGTGGGTAAAGGCACAAAGACACCCAGCGTGCCAATGTACACAGTT	47231	
Qy	5643	aaacgaagtaacgcaactgttgggtcttggtaatgctgtgtaataacacgtgcagcga	5702	

Db 47232 AAACGAAGTACGCAACTTGTGGGTCTTGGTAATGCTGGTAAATGATAACGCTGACGGCAA 47291  
Qy 5703 tcaggttaaacattgcccgcacatacaaaaagaccacaaattcaggtttcatcatcataccgacac 5762  
Db 47292 TCAGGTAAACATTCGCCGACATCAAAAAGACCCAAATTCAGGTTCAATCATTAACCCGAC 47351  
Qy 5763 tgcataaagcaggacacgggtacttggcgttaaggttaataaacgataccgaaaaacttgc 5822  
Db 47352 TGTATCAAAAGCAGCAGCGTACTTGGCGGTAAAGGTAAATAACGATACCGAAACACTTGC 47411  
Qy 5823 caactggtgtgataagtggtgataaagacgacgacactaacgagcgtatttaagcaa 5882  
Db 47412 CACTGTGTGTATACAGTGGCGTGGTAAAGACGCGAAGCTAACGCGATTTAAGCAA 47471  
Qy 5883 tgtttgggtcaaaaccccaaaagatggcagcaaaaaagccctgtcgcacctataaagc 5942  
Db 47472 TGTTTGGGTCAAAACCCCAAAAGATGGCAGCAAAAAGCCCTGCTCGCACATTATAACGC 47531  
Qy 5943 cgcaggtcaagacaactatttgacacaacacccgcagaagccattgacagaataaataga 6002  
Db 47532 CGCAGGTCAGACCAACTATTTGACCAACAAACCCCGCAGAACCCATTCAGACATTAATGA 47591  
Qy 6003 aceagtatccgcttcttcacatgcaacgatggcaatcaagagccctgtgtacaaggcg 6062  
Db 47592 ACNAGGTATCCGCTTCTCCATGTCAACGATGCAATCAAGAGCCGTGGTACAAGGGCG 47651  
Qy 6063 taacggcattgactcaagtcctcaggaagaacactcagtggtgagatagtttccagggcaa 6122  
Db 47652 TAACGGCATTTGATTTCAAGTCTGCTGGCAAGCATTCAGTGGCGATAGGTTTCCAGGCCAA 47711  
Qy 6123 ggcagatggtgaagccgcctgtgcaataggcagacaaaccccaagcaggcaaccaatccat 6182  
Db 47712 GGCAGATGTTGAAGCCGCGCTGCCATAGGCAGACAAACCCAGCAGGCAATCCAT 47771  
Qy 6183 cgcacatggtgataacgcagaaacccgcagggatcaatcatcgccatcggttacaggcaa 6242  
Db 47772 CGCCATCGGTGATAAAGCAACGCAACGCCAGCGGCGATCAATCATCGCCATCGGTACAGGCA 47831  
Qy 6243 tgtggttaacaggttaagcaactctgtgccaatcgccgaccccaagcaactgttaaggctgataa 6302  
Db 47832 TGTGCTAGCAGGTAAAGCACTCTGTGCTGCAATCGCGACCCCAAGCACTGTTAAGGCTGATAA 47891  
Qy 6303 cagttacagtggtggttaataaacacagtttatcgatgccactcaacagatctcttgg 6362  
Db 47892 CAGTTACAGTGTGGGTAAATCAACACCAAGTTTACCAGATGCCACTTCAGCCGATGCTTTGG 47951  
Qy 6363 tgtgggcaataacataccctgaccggaagtaactcgtggttgccttgggttcaaacctcgc 6422  
Db 47952 TGTGGGCAATAAACATACCGTGACCGAAAGTAACCTCGGTTGCTTAGGTTCAAACTCTGC 48011  
Qy 6423 catcagtcagggcacacacgcaggacacacagcccaaaaatctgacggcacacaggtac 6482  
Db 48012 CATCAGTGCAGGCACACACGACGACACAAAGCCAAAATCTGACGGGCACACAGGTAC 48071  
Qy 6483 aaccacacacaggttcaacaggtacggttaaggtttgcttggaacaaacggttgg 6542  
Db 48072 AACCACACACAGGTGCCACAGGTACGGTTAAAGGCTTTGCTGGACAAACGCGGTTGG 48131  
Qy 6543 tgcggtctccgtggtgctcaggtgtgtgaacgctgtatccaaaatgtgagcaggtga 6602  
Db 48132 TGCAGTCTCCGTGGGTGCTCAGGTGTGTAAGCGCGTATCCAAAATGTGACAGGTGA 48191  
Qy 6603 ggtcagttcaacacagcaggtatcggtcaatggttagccagttgtacaagccacccaag 6662  
Db 48192 GGTAGTGCACACAGCAGCGATGGGTCAATGTTAGCCAGTTGTACAAGCCACCAAG 48251  
Qy 6663 cattgccacacacaaatgagcttgaccatcgatccacacaaacaaataaagcaca 6722  
Db 48252 CATTCGCCAACGCAACCAATGAGTTGACCATCGTATCCACCAAAACGAAAATAAGGCAA 48311  
Qy 6723 tgcagggatttcaatcagcaggtggtggtggtcctcagtcacacagcctacatcctgag 6782  
Db 48312 TGCAGGGATTTTCATCAGCGATGGCGATGGCGTCCATGTCACACAGCCTACATTCCTCGCAG 48371

Qy 6783 atccatggttacgggggtattggtccaccacacacaggtcaaggtgcggtggtgagact 6842  
Db 48372 ATCCATGTTTACCGGGGTATTGTCACACCCACACAGCGTCAAGGTGCGGTGCGAGTGGGACT 48431  
Qy 6843 gtcgaagctgtcggataaattggtcaatgggtatttaaaaaatcaatggttcagcgacaccca 6902  
Db 48432 GTCGAAGCTGTCGGATAATGTCATGGTCAATGGGTATTAAATCAATGTTTCAGCCGATACCCA 48491  
Qy 6903 aggcacatgtagggggcgccaggtggtcaggttttccacttt 6942  
Db 48492 AGCCATGTAGGGGACGACGTTGTCGAGGTTTTCACTTT 48531  
RESULT 4  
AX079918 AX079918 6141 bp DNA PAT 22-FEB-2001  
DEFINITION Sequence 6 from Patent WO0107619.  
ACCESSION AX079918  
VERSION AX079918.1 GI:13159439  
KEYWORDS Moraxella catarrhalis.  
SOURCE Moraxella catarrhalis.  
ORGANISM Moraxella catarrhalis  
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;  
REFERENCE 1 (bases 1 to 6141)  
AUTHORS Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.  
TITLE Recombinant high molecular weight major outer membrane protein of moraxella  
JOURNAL Patent: WO 0107619-A 6 01-FEB-2001;  
CONNAUGHT LABORATORIES LIMITED (CA)  
FEATURES  
source Location/Qualifiers  
1. 6141  
/organism="Moraxella catarrhalis"  
/db\_xref="taxon:480"  
BASE COUNT 2012 a 1403 c 1399 g 1327 t  
ORIGIN  
Query Match 37.3%; Score 2590; DB 10; Length 6141;  
Best Local Similarity 66.9%; Pred. No. 0;  
Matches 4802; Conservative 0; Mismatches 1105; Indels 1269; Gaps 28;  
Qy 1 atgaatcacatataaaagtcatctttaacaaagccacagggacacatttatggcgtggcga 60  
Db 1 ATGAATCACATCTATAAAGTCATCTTTAAACAAGCCACAGGCACATTTATGGCAGTGGCA 60  
Qy 61 gaggtcgcaaaatccacacagggagggagtagcagtagcagcagcagcaggtgggcagc 120  
Db 61 GAGTACGCCAAATCCACAGACACGGGGGGGTACTGTGTACAGGGCAAGTTGGCAGT 120  
Qy 121 tctcctgtcatccgctgactcgtgtgtccacgctcgctatactctagtgtatcggtgcgacg 180  
Db 121 GTATGCACCTCTGAGCTTTGCCGCTATTGCGGCTCGCTGCTGCTGATCGTGCAACG 180  
Qy 181 ctcaatggcagtgcttatgctcaaaa-----ataatagcaagatcgcatctgtt----- 228  
Db 181 CTCAGTGGCAGTGTCTATGCTCAAAAAAAGATACCAACATATCGCAATTGGTCAACAA 240  
Qy 229 ---accacagggcaacaatgacatgacctcggtgctagcaatgaagcatccattgctattggt 285  
Db 241 AACCAGCCAAAGCGCTCAGGCATCTGCCAAGCGGAGGGGTATCGAGCCATTGCTATTGTT 300  
Qy 286 agtcttgcataagggcacatgccaatcaagctattgtctatcggtggtagcaaacacagatcct 345  
Db 301 GAAAAATGCTAACACACAGGGCGGTCAAGGCATCGCCATCGGTAGTAGTAA---TAAACT 357  
Qy 346 cgttaatcaagcggcctaatacagaagcaggttccacgcccacaaaggttaagagtcacgcgc 405  
Db 358 GTCAATGGGAAGCAGTGTGGATACATAGGTACCGATGCTACCGGTCAAGAGTCCATCGCC 417  
Qy 406 atcgggtggtgtagtactggtggtggtggtggtggtggtggtggtggtggtggtggtggt 465

Db 418 ATCGTGGTGATGTAAAGGCTAGTGGTGATGCCCTCGATTGCCATCGGTAGTGATGACTTA 477  
Qy 466 tatttgatgaatagcactaactctaaataatccaaatgg---tcttcttagcactctt 522  
Db 478 CATTTGCTTGATCAGCATGCTAATCTAAACATCCGAAAGGTACTCTGATTAAAGTACTT 537  
Qy 523 attcaaaaccatacagattattacgccaaataacgagactcaaaatggtctca---gaaatat 579  
Db 538 ATTAACGGCCATCGCTATTATAAGAAATACGAAGCTCAAAAGGATATATGATGTAAATAT 597  
Qy 580 agagcacagcagcagaagacagccagctactgctgagtgagccatggccatgcaatgcaaaag 639  
Db 598 AGACGCACACCCGACAGCGGACAGCCAGTACTGTCAGTGGAGCCATCTCATATGACAG 657  
Qy 640 ggtcattttgccacgaccttggtagcaggtcactaacagctgaagcgaactatctctggca 699  
Db 658 GGTCAATTTTCCAAAGCGCTTGTGTACACGGGCAACAGCTAAAGTGCCTATTCTCTGGCA 717  
Qy 700 gtaggcttaccgccaagccgaaagagataatacactcgctattggttcttaataatgcacaa 759  
Db 718 GTGGGTCTTCCGCCACAGCGGAGGCCAATCTACAATCGCTATTGGTTCGTGATGCAACA 777  
Qy 760 gctatcaattatgagcactagcccttggtagcagatactcagctgattgattggattacggt 819  
Db 778 TCTAGCTCGTGGAGCGCATAGCCCTTGGTGGAGGTACTCTGCTCAGCTACAGGGCAGT 837  
Qy 820 attgcccattggttgcctcagactccttaataataataataata---ataataaa 876  
Db 838 ATTGCCCTAGGTCAAGGTTCTGTGTCTACTCAGAGTGATATAATTTCTAGACCGGCCAT 897  
Qy 877 gctatgtaccagaaggtatggttcaaacataaaatcgtctaaagccaccgccaatggt 936  
Db 898 ACACCAANTACCCAGGCACCTAGACCCCAAGTTTCAAGCCACCAATAATACGAAGGCGGGT 957  
Qy 937 ttatttccattggttagcactatacaagctcaaatcaatcgaatgctgaggttat 996  
Db 958 CCATTTCCTATGGTAGTAATCTATCAACAGTAAATCATCAATGTCGGTGCAGGTGTT 1017  
Qy 997 gaggataccgctggttcaatgtggcacagctaaagcgtgagaaatctggcta---ag 1053  
Db 1018 AATAAACCAGTACGGTCAATGTGGCACAGCTAGAACGGTGGTGAAGTGGGCTAAGGAG 1077  
Qy 1054 cgtcaattactttaaaggtgatgataacggtactggttgaagaaacactggcgag 1113  
Db 1078 CGTAGAATTAATTTACGGGTGATGATACAGTACTGACGTAAATAATAGTTTGGATAAT 1137  
Qy 1114 actttaaaccattaaagtggtgagaccgaagcgacaaagctaaaccgataataataacatt 1173  
Db 1138 ACTTTAACTATTAAAGTGGTG---CAGAGACCAACCCATTAAACCGATATAATATCGGT 1194  
Qy 1174 ggtgtgtaacagataataacttggtgtaaaagttaaacttgctaaaaacctaagcgtt 1233  
Db 1195 GTGTAAAGAGGGCTGATAATAGTGGTCTGAAAGTTAAACTTGTCTAAACCTTTAAACAAT 1254  
Qy 1234 ctggaacaggttagcccaaaacctaaccgagcgagagaagttacgtag---gtagt 1290  
Db 1255 CTACTGAGGTGATACACTACATTAATGCCCACACACAGTAAAGGTAGGTAGT 1314  
Qy 1291 ggtataaacccgctgagctacaaaagcgttgggttaaaccctttaccc-----caacaaca 1344  
Db 1315 AGTAGTACTACAGCTGAATTTATGAGTGTATGATGTTTAACTTTACCCAGCCCAATACAGGC 1374  
Qy 1345 aatgcaagcacagacaaaacccgtctatgagcactgagtggttgaatttactgataatct 1404  
Db 1375 AGTCAAGCACACGAACAAACCGCTATGGGCTTAATGGGTTGAAGTTTACTTAATAATGCA 1434  
Qy 1405 aatcggcacttga---gatactactcgtatcaccaagaataaattggttttagcaat 1461  
Db 1435 GAACAACACAGCAATTCGGCACTACTCGGTATTACCAGAGATAAAATTTGGCTTTGCTCGA 1494  
Qy 1462 aaagctggtcagttgataaacaacacttatcttgaataaagacaaagctaaaaag----- 1516  
Db 1495 GATGGTGATGTTGATGAAAAACAAGCACCATATTGATGAAAAACAACACTTAAAGTGGGT 1554

Qy 1517 -----tt 1518  
Db 1555 AGTGTTCATATTACCATAGACAATGGCATTTGATGCAGGTAAATAAAAAAGATCAGTAATCTT 1614  
Qy 1519 ggaacaagcagccctaaacaacaggttgcttgactgttaataaacaccattggtggttagcaat 1578  
Db 1615 GCCAAAGGTAGCAGTGTACGATGCGGTACCATCAACAGCTCAAAAGCGCCGAAGCCT 1674  
Qy 1579 aaacaataccaagtcggtgctgatggcattaaatttggcagtggaatgttaattatca 1638  
Db 1675 ACITTAACCCAGCGCTGCGCATCTGTCACACCTACTGAAATATCAGTTGATGCTTAAG 1734  
Qy 1639 aatgcgcgaataatgcgcactactcgtattaccogaagagaaattggtcttgcgtgact 1698  
Db 1735 AGTGGCAATTTACCGCCCAACTTACACATTTGGCTGAAACACACCGAGCTTAACAGT 1794  
Qy 1699 gatggtaaagtgtgataa----- 1715  
Db 1795 GATGGCACTAGTGATATAATTTAGGTAAAGGTAGTGGTACGAACAATAGCTTATAGTTAAC 1854  
Qy 1716 -----aaagtcaccatatttgataaaaaaacaacttcaagtggtggtt 1761  
Db 1855 GCCCAACATTTGGCAAGCTATCTAAATGAAGTCAATCGAACGGCTGACAGTCTTACAA 1914  
Qy 1762 aaattaccaaaacagtggtcattaatgcaggtgatcaaaagatcagtaattgttaagat 1821  
Db 1915 AGCTTTACCGTTAAAGAAGAAGACGATGACGCCAACGCTATCACCGTGGCTAAAGAT 1974  
Qy 1822 gcaacgagcagatccgagtcagtcactataaacaggttataaacagtcacaagaagcgc 1881  
Db 1975 ACGACAAAAAATCGCGCGAGTCAGCATCTTAAATCTCAAAAGTAAACACGGTCTAACG 2034  
Qy 1882 gacggtgcccataca-----agcttctcttctcgtgatgaaaggttcagga 1929  
Db 2035 GTTGTACCAAAAAAGATGTCAGGTTACCTTTGGGCTAGCCAGATACGGTCTGACC 2094  
Qy 1930 ttacgattagtaacttgatttcttaattgtaataaccccaataacctttgagaccatcacc 1989  
Db 2095 ATTGGCAAAAGCACCCCTAAACACACGATGGCTGACTGTTAAAGATACCAACGAACAATC 2154  
Qy 1990 ttgcaagtgaaaacgcatcagtatcagcaatgacatagccaaagtaaaagtcaaaag-- 2047  
Db 2155 CAAGTCGGCTGTAATGGCATTAAATTTACTAATGTGAATGTAGTAATCCAGGTACTGGC 2214  
Qy 2048 -ttggtattgaccaatcaatggctcaccagcctgaagctgacctgggtgagcgtagcgtataa 2106  
Db 2215 ATTGCAANTACCGCTCCGATTACCCAGAGATAAAATTTGGCTTTGGTCTGTGATGGTGCA 2274  
Qy 2107 gatggttaaaactcaattggttattgagcaag----- 2137  
Db 2275 GTTGATACAAACAACCTTATCTTTGATCAGACAAGCTACAAGTTGGCAATGTTAAGATT 2334  
Qy 2138 ----tggttagcgttaacgacacaccccaataatcattagagattgctcccaacactcct 2193  
Db 2335 ACCAACACTGGCATTAAACAGGTGTTAAAGCCATACAGGGCTGTCCCAACACTGGCT 2394  
Qy 2194 agcattaccaatgaggtgagtcagcaccacagagagcaatacaatacaccagcagac 2253  
Db 2395 AGCATTCCCGATCAAGTAGC---CGCAACATACAACTGGGCAATACAATCCA---AGAC 2448  
Qy 2254 gaagcaaaatccaaagcgcagctatcggtgatatataataataacagcgtttaaactaaaa 2313  
Db 2449 AAAGACAAATCCACGCTGCCAGCATTAATGATATATTAATACAGGCTTTAAACCTAAA 2508  
Qy 2314 aataatgcaactcgttgcttctcacttataaaccttataaaccttgcattatcgatggc 2373  
Db 2509 AATAAACAACCCCATAGCTTTGTCTCCACTATGACATTTGTGTGACTTTGCCAATGGC 2568  
Qy 2374 aatgccaccacgcgtgaaggtaacttacgatgaacccaatcaaacagtaaaagttaacttat 2433  
Db 2569 AATGCCACCCGCCACAGTAACCCATGATACCCGTAAACAAACCACTAAAGTGGTATAT 2628

QY 2434 gatgcaatgtgtagaagaaacacattgaactcacaggcgataatggcgaagacacacaaa 2493  
 Db 2629 GATGTGAATGTCGATGATACACCAATTCATTCACAGCGACTGATGACAA---TAAAAA 2685  
 QY 2494 attggcgtaaaacacacacactgacccacacaaatgtaagtgttaagaaacac--- 2547  
 Db 2686 CTGTGGCGTCAAAACACCAACCTGAACAAACAGTGTAAATGATGTAATACAGCAACTAAC 2745  
 QY 2548 ---aactttagtaacacagataacgtagccctgttaacgcccacaaagacatcgccgaata 2604  
 Db 2746 TTTAATGTTAACTTACTGATGAAGATGCCCTGTGTTAACGCCAAAGACATCGCCGAAAT 2805  
 QY 2605 ctaaacacctagccaaagaaattcacacccacaaagacacagacagacacccctacaa 2664  
 Db 2806 CTAACACACCTAGCCAAAGGAAATTCACACCACCAAGGCACAGACACCGCCCTACAA 2865  
 QY 2665 acctttaaagtcacaaagacggtgca-----actgatgacgaaacacatcacctggtg 2718  
 Db 2866 ACCTTTACCGTTTAAAGGTAGATGAAATATATATATGCTGATGACGCCAACGCCATCACC 2925  
 QY 2719 aaagatggttacacaaacggaagacgctcaacactctaaactcaaaactcaaaaggtgaaacgg 2778  
 Db 2926 GTGGGTCAAAAGAACGCAAAATATCAAGTCAACACCTCAACACTCAAGAGGTGAAACGGT 2985  
 QY 2779 ctaacgggtgctaccaataaagatgacggttaccttggcattacacaccccaagcgg 2838  
 Db 2986 CTTAATATTTAAACCCGACAAATGTTACGTTACCTTTGGCATTAACACACCAAGCGT 3045  
 QY 2839 cttaaagccgacgacgacacacactctaaacaaagatgggttctctattaaacaccccg 2898  
 Db 3046 CTTAAAGCGGCAAAAGCAC---CTAAAGACGGTGGTGTGCTATTAAACACCCACT 3102  
 QY 2899 agtaacgaacaaatccaagtcggtgctgtagcgtagtaagtttgcgaagt---tgataag 2955  
 Db 3103 GGTAGCGCAAAATTCGAAGTCGGTGTGATGGCGTGAAGTTTGCCAAAGGTTAATATAT 3162  
 QY 2956 ggtatcaacgactgattgtaggcacagcgtatcaccaagcgtatcaccaagacataatggcttt 3015  
 Db 3163 GGTCTGTAGTGTGCTGCAATTTGATGGCACAACCTCGCATTTACAGAGATGAATTTGCTTT 3222  
 QY 3016 actgggggtaagtctgactgactgatacacaacacccacacacacacacacacacacac 3075  
 Db 3223 ACTGGGACTAATGGCTCCTTTGATTAAGCAAAACCCACCTTAGCAAGAC----- 3273  
 QY 3076 gtgggtgaagtgaattaccaacactggcattacacgaggtggtggtggtggtggtggtggt 3135  
 Db 3274 -----GGCATTAACGAGGTGTTAAATAAGATTTACCAAC 3306  
 QY 3136 attcaatcaggtgatattaccacaaacacagcaatgactgtagcagcggtcggtttat 3195  
 Db 3307 APTCAATCAGGTGAGATTGCCCAAAACAGCCATGATGCTGTGACAGCGCGCAAGATTAT 3366  
 QY 3196 gatttataaaacgacttgaagcaaaatcacagctgctgtataaaacagacacaaactca 3255  
 Db 3367 GATTTAATAACCGAATTGAAACAAATACAGAGTACTGCGCAAAACAGCAAAACTCA 3426  
 QY 3256 ttacacgaattctcagtagcagatgaacaaaggttaatoactttacggttagtaacccctac 3315  
 Db 3427 TTTACAGCAATTTCTAGTAGCAGATGAACAAGGTAATACTTTACGGTTAGTAACCCCTTAC 3486  
 QY 3316 tccagttatgacacactcaaacctctgactgctacacctttgacaggtgaaaaacggcatt 3375  
 Db 3487 TCCAGTTATGACACCTTCAAGACCTTCTGATGTCATCACCTTTGACGGTGAACCGCAT 3546  
 QY 3376 accacaaaggtataaaggtgtggtggtggtggtggtggtggtggtggtggtggtggtggt 3435  
 Db 3547 ACCACCAAGGTAAATAAAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 3606  
 QY 3436 acgcctaaactgacggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 3495  
 Db 3607 ACGCCCTAAGCTACCGGTGGTGAATTAATGCGCAAGGCAATTTGTCATTTGACAGCCAAAT 3666  
 QY 3496 ggtaaaatatacatcacaggaactaacactctagctaatgtttaccaatga-----t 3549

Db 3667 GGTCAAAATACCATCACAGSACTTACCAACTCTAGCTTAATGTTTACCATGATAAAGGT 3726  
 QY 3550 ggtgcaggacacgacactaagcgaaggttgcgaatgacacacgacacacacacacacacac 3609  
 Db 3727 AGCGTACGCCACCAAGAACAGGGCAATATATCAAAAGACGAAGACAAACCCCGTCCGCC 3786  
 QY 3610 agcatgggtgtagtgcataaacgaggtcttaacttgaagggcaatgggtgaagcgggtgac 3669  
 Db 3787 AGCATGTTGTTGATGTTAAGCGCAGGCTTTAACTTCAAGGCAATGTTGAAGCGGTTGAC 3846  
 QY 3670 ttgtctccacttatgacactgttgcactttatgcataaggaatgcacacacacacacacac 3729  
 Db 3847 TTTGCTCCACTTATGACACCGTCAACTTTGCGGATGGCAATGCCACCGCTGTAAGGTG 3906  
 QY 3730 acctatgatac 3789  
 Db 3907 ACCTATGATGAC 3966  
 QY 3790 accttgaagtgcacaaatgcataaaaaacttggcgctcaaaacacacacacacacacac 3849  
 Db 3967 ACCATTGAAGTTAAA---GATAAAAAACTTTGGCGTAAAAAACACACACTTTGACCGACT 4023  
 QY 3850 agtctaatgtaatgca---accataatttagtgcgcgcatggcgatgccccttgtttaa 3906  
 Db 4024 GGCACAGGTGCTAATAATTTGCCCTTAAGCAATCAAGCTACTGGCGATGCGCTTGTCAAG 4083  
 QY 3907 gccagtgatatcgccac 3966  
 Db 4084 GCCAGTGATATCGTTGCTCATCTTAACACACCTTATCTGCGGACATCCAAACTGCCAAAGG 4143  
 QY 3967 gcaagcgaagcaagcagctcagcagctatgtgctgctgctgctgctgctgctgctgct 4026  
 Db 4144 GCAAGCCAAGCGAACCACTCAGCAGCTATGTGATGCTGATGCAATAGTGCATCTAT 4203  
 QY 4027 gacagtacgcataagaaatgactatcaagtcaatgcaaggttcaagtggacacacacacac 4086  
 Db 4204 GACAGTACCGATAACAGTACTATCAAGCAAAATGATGACAGCTGTGATTAACCAACAA 4263  
 QY 4087 gaagtgcacaaagacacactggtgcgcacacacacacacacacacacacacacacacac 4146  
 Db 4264 GAAAGTGGCAAAAGCAAACTGGTGCCTCCACAGCCCAACCCAGATGGCAGCATGGCTCA 4323  
 QY 4147 atgaatgtcaaatcagctattacacaaagcaagtataatgctgctgctgctgctgctgct 4206  
 Db 4324 ATGAATGTCAATCAGTCAATTAACAAAGCAAGTAAATGATGCCAATTAACCAAGCGC 4383  
 QY 4207 atcaatgaagac 4266  
 Db 4384 ATCAATGAAGACACACCGCTTTGTTAAGGACTTGAAGACCGCTTCTTGATTAACCAAC 4443  
 QY 4267 aaaaacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 4326  
 Db 4444 AAAAAACCGCGAGTACTGCTGGGTGATTTAATGCGTGGTGGTGGTGGTGGTGGTGGT 4503  
 QY 4327 gcaagggatcacaggac 4386  
 Db 4504 GCAGGGGATACAGGCACACCGCTAAAAAATCTGGCGAGACTTTGACCATCAAGAGTGGG 4563  
 QY 4387 caaaagac 4446  
 Db 4564 CAAACAGACACCAANTANGCTAACCCGATATATACATCGGTGTGGTACGAGTACTGATGSC 4623  
 QY 4447 ttcaactgcaaaacttgcacaaagacacacacacacacacacacacacacacacacacac 4506  
 Db 4624 TTCACGTCAAACTTGCAAAGACCTTAACCAATCTTTAACAGCGTTTAATGACAGTGGCACC 4683  
 QY 4507 agaatgataaaaaagcactctctttagacgcaacacacacacacacacacacacacacac 4566  
 Db 4684 AAAATGATGACAAAGCGGTGCTCTTTGTAGACTCAAGCGGTCAAGCCCAAGCAACACAC 4743  
 QY 4567 cctgtgtgaagtgcacaaatggcctggacccctgggtggcaaacacacacacacacacacac 4626

Db 4744 CCTGTGCTAAGTGCCTAATGGCTGCTGGACCTGGGTGGC----- 4779  
Qy 4627 gctgttgatgataacgagtcggtgaaactttaagcagtttaagcagtttgccaaacggtc 4686  
Db 4780 ----- 4779  
Qy 4687 acaacctaacaacaaaagtaactcaggctgctgctaccccttggttaacogatgcc 4746  
Db 4780 ----- 4779  
Qy 4747 aatggcaagcccatcaatgacacgagtggaagcccaaaagccatcaagggcgccgat 4806  
Db 4780 ----- 4779  
Qy 4807 ggttaatactatcacgccaacgccaacgctgacacgctgacaaagatggcaagcccatc 4866  
Db 4780 ----- 4779  
Qy 4867 accgatcggaacaaacttgccaatctggcagctcatggcaaaccccttgatgacaggtcat 4926  
Db 4780 ----- 4779  
Qy 4927 caagtgtggaagccttagcggaactcagatgccatcaccccttaacacatcaagtcc 4986  
Db 4780 ----- 4779  
Qy 4987 actttgccaaaattgacacacacaaacacaggttaatgccaatgagggcaagcccaagt 5046  
Db 4780 ----- 4779  
Qy 5047 ctgcccagcctatcagcagcacagcaaaagtaactgctccagtgctcaaaagatgctgaat 5106  
Db 4780 ----- 4779  
Qy 5107 gtaggctttaacttgcagaccaaatacaatacaagtggactttgtcaaaagcctatgatacc 5166  
Db 4780 ----- 4779  
Qy 5167 gtcaactttgtcaatggtataggtgcgcgacatcaacaagcgtgctagtgctgtagggcag 5226  
Db 4780 ----- 4779  
Qy 5227 atgagtaacatcaacgctcaacacgcgcttagcagcgaccgatgatgagtgcaatgctctt 5286  
Db 4780 ----- 4779  
Qy 5287 atcaagcgaagatggttaagttctacaagcagacacctcatgccaaagcgtcacta 5346  
Db 4780 ----- 4779  
Qy 5347 aaagcaggcaaatcagccagtgatgccaactccaactggtctaaagccttgttaacccc 5406  
Db 4780 ----- 4779  
Qy 5407 aatgctgtaaaagcagtagcagtgatgagtgctctttaacttaactatcaaaagcggta 5466  
Db 4780 ----- 4779  
Qy 5467 tttaaatocaaagatggtataactactaccacagtaagctctgatggcatcatgtaacaa 5526  
Db 4780 ----- 4779  
Qy 5527 ggcaaaagataacagcagcatcacccctaaagcaagatggctgaatgtaggcggtaagtc 5586  
Db 4780 -----AAGTGC 4785  
Qy 5587 atcagaatgtggttaaaagcacaacaaagacacgacgctgccaatgtacaaagctttaaac 5646  
Db 4786 ATCAGTAATGTGGCAAGGCACAAAAGATACCGACGCTGCCAATGTACAAAGCTTTAAAC 4845  
Qy 5647 gaagtacgaactgttggctcttggttaagtctggttaagtataacgctgacggcgaatcag 5706  
Db 4846 GAAGTACGCAACTGTGTGGTCTTGGTAAATGCTGTAATGATACAGCTGACGGCAATCAG 4905

Qy 5707 gtaaacattccgacatcaaaaaagaccccaattcaggttcatctatcctaaacgcactgtc 5766  
Db 4906 GTAACATTCCGACATCAAAAAGACCCAAATTCAGTTCATCATCTAATCCGACACTGTC 4965  
Qy 5767 atcaaacagcagcgacttgcggttaaggttaataacgataccgaaaaacttgcacat 5826  
Db 4966 ATCAAAAGCAGCAGCTACTTGGCGGTAAGGTAAATACGATACCGAAAACTTGCACACT 5025  
Qy 5827 ggtggtgtacaagtggtggtataaagacggaagcgaactaacgctgaagcgaatggtt 5886  
Db 5026 GGTGTTATACAAAGTGGCGGTGGATAAAGACGGCAACGCTAACGGCGATTTAAGCAATGTT 5085  
Qy 5887 tgggtcaaaaccccaaaagatggcagcaaaaaagccctgctgcacattataacgcgcga 5946  
Db 5086 TGGGTCAAAACCCCAAAAGATGGCAGCAAAAAGCCCTGCTCGCCACTTATAACGCCGCA 5145  
Qy 5947 ggtcagacccaactatttgaccaacaaccccgacagaacccattgacagaataataagaaca 6006  
Db 5146 GGTGAGACCAACTATTATGACCAACACCCCGCAGAAGCCATTGACAGATAAATGAACAA 5205  
Qy 6007 ggtatccgcttctccatgtcaacgatggcaatcaagcctgtgtacaaagcgtaac 6066  
Db 5206 GGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCCTGTGGTACAAAGGCGTAAC 5265  
Qy 6067 ggcattgactcaagtgcctcaggcaagcactcagtgcgataggtttccagggcgaagca 6126  
Db 5266 GGCATTGACTCAAGTGCCTCAGGCAAGCACTCACTGGCGATAGTTCAGGCGCAAGGCA 5325  
Qy 6127 gatgtgaagcgcgcttgccatagggcagacaaccccaagcaggcaacccatccatcgcc 6186  
Db 5326 GATGGTGAAGCGCGCTTGGCCATAGGCAGACAAACCCCAAGCAGCAACCAATCATCGCC 5385  
Qy 6187 atcgtgtatacgcacacgaagcgaagcgaatccatccatcgctgtacaggaatgtg 6246  
Db 5386 ATCGGTGATACGCAACGACCGCGGCGATCAATCCATCGCCATCGGTACAGGCAATGTG 5445  
Qy 6247 gtaacaggtgaagcactctggtgccatcggcagcccaagcactgttaaggtctgataacagt 6306  
Db 5446 GTACAGGTAAAGCACTCTGTGGTCCATCGGCAGCCCAAGCACTGTTAAGGCTGATAACAGT 5505  
Qy 6307 tacaagtgggtaataacaacacagtttatcgatgccactcagacccgatgtcttgggtg 6366  
Db 5506 TACAGTGTGGTAAATAACAACCACTTTACCGATGCCACTCAAAACCGATGCTTTTGGGTG 5565  
Qy 6367 ggcaataacatcacctgacgcgaagtaactcgttgccttaggttcaaaactctgcccac 6426  
Db 5566 GGCATATACATCACCGTGCAGCAAGTAACTCGCTTGGCTTAGGTTCAAACTCTGCCATC 5625  
Qy 6427 agtgaggcacacacgagcagcacacaagcgaacaaaaacttgacgagcagcaggttacaacc 6486  
Db 5626 AGTGCAGGCACACACGCGAGGCACACAAGCCAAAAAATCTGACGGCACAGCAGGTACAACC 5685  
Qy 6487 accacagagtgcaacaggtacaggttaaaagcttctggtgacaaaaagcggttgggtgcg 6546  
Db 5686 ACCACAGAGTGCACACGCTTACCGTTAAAGGCTTTGCTGGACAAACGCGGTTGGGTGCG 5745  
Qy 6547 gctccctgggtgctcaggtgctgaacgcgctatccaaaaatgtggcagcaggtgaggtc 6606  
Db 5746 GTCTCCGTGGGTGCTCAGGTGCTGACCGCGTATCCAAAATGTGGCAGCAGGTGAGGTC 5805  
Qy 6607 agtgcaccagcagcagtggtggtcaatggtagccagttgtacaaagccaccccaagggcatt 6666  
Db 5806 AGTGCCACACGACCGCGTCCGTTCAATGGTAGCCAGTTGTACAAAAGCCACCAAGCATTT 5865  
Qy 6667 gccaagcacaacaaatgagcttgacatcgatccacacaaaaagcaaaaaataaagcgaatgca 6726  
Db 5866 GCCAACGCAACCAATAGCTTGAACCATCGTATCCACCAAAACGAAAAATAAGGCGCAATGCA 5925  
Qy 6727 gggatttcacagcagtggtggtggtcctcagtcgaacagcctacattcctcggcagatcc 6786  
Db 5926 GGGATTTCATCAGCGATGGCGATGGCGTCCATGCCAANAGCCTACATTCCTTGGCAGATCC 5985







1517 -----tt 1518  
Db 2096 AGTGTTCGAATTACCATAGACAAATGCAATGTATGATGCGAGGTAAATAAAGATCAGTAATCTT 2155  
1519 ggcacagcacccataaacacggtggtgactgttaataaacacacattggtgtagcaat 1578  
Db 2156 GCCAAGGTAGCAGTGTACGATGCGGTATACCATGCAACAGCTCAAGCCGCAAGCCT 2215  
1579 aaacaaatccaagtgtgctgtaggcatataatattgcccagtgtagtgaattatgaatca 1638  
Db 2216 ACPTTAAAGCGCAGCGCTGGCATCAGTGTACACACCTACTGAAATATCAGTTGATGCTAAG 2275  
1639 aatgcccgaataatcgccactactcgtattaccgaagagaaattggtccttgcgtatgt 1698  
Db 2276 AGTGGCAATGTTACCGCCCAACTTACACATTTGGCGTGAACACACCGAGCTTAACAGT 2335  
1699 gatggtaaagttgataa----- 1715  
Db 2336 GATGGCACTAGTATAAATTAGTGTAAAGGTAGTGGTAGCGAACAATAGCTTAGTTACC 2395  
1716 -----aaagtcaccataatttggataaaaaaacacttcaagtgggtggtt 1761  
Db 2396 GCCGAACAATTTGGCAAGTATCTAAATGAATCAATCGAAGCGCTCACAGTGTCTTACAA 2455  
1762 aaattaccacagacagtgagcattatgacagtgatcaaaagatcagtaagttaagat 1821  
Db 2456 AGCTTTACCGTTAAAGAAAGACGATGATGACGCCAACGCTATCACCGTGGCTAAAGAT 2515  
1822 gcaacggaagatccgtagtcactctataaacagcttataaacagcttataaacagcgc 1881  
Db 2516 ACACAAAAAATGCGCGCGAGTCAGCATCTTAAACTCAAGGTAAAAACGGTCTTAACG 2575  
1882 gacggtgcctacaa-----agcttctctattcgtgatgaaagagtcaggaa 1929  
Db 2576 GTTGTCTACCAAAAAAGATGGTACGGTTACCTTTGGGCTTAGCGCAAGATAGCGGTCTGACC 2635  
1930 tttaacattagtaactgtattcttaagtataatccccaataaccitttgagaccatcac 1989  
Db 2636 ATTGGCAAAAGACCCCTAAACAACGATGGCTTGACTGTGTAAGATACCAACGAACAAATC 2695  
1990 ttgacaggtgaaacggtacagtaacagcaatgacatgacgaaggaagtaagtaag-- 2047  
Db 2696 CAAGTCGGTGTATGSCATTAATTTACTAATGTGAATGGTAGTAATCCAGGTACTGGC 2755  
2048 -ttggtattgacccaatgaatggtctccacacgccttaagctgacctgacctgtagcgtataa 2106  
Db 2756 ATTGCAAAATACCGCTGCGATTTACAGAGATAAAATTTGGCTTGTGCTGTGATGCTGCA 2815  
2107 gatggtaaaaactcaattggttattgagcaag----- 2137  
Db 2816 GTTGATACAAACAACCTTATCTTGTATCAAGACAAGCTACAAGTTGGCAATGTTAAGATT 2875  
2138 ----tggtagcggttaacgacacacacacacacatcatctagaggattgtcccaacactgcct 2193  
Db 2876 ACCAACACTGGCATTAACGACAGGTGTTAAGCCATCACAGGGGTGTCCCAACACTGCCT 2935  
2194 agcattaccaatgcaggtggtgctacgcacacacacacacacacacacacacacacacac 2953  
Db 2936 AGCATTTCCCGCATCAAAAGTAGC---CGCAACATAGAACTGGGCAATACAAATCCA---AGAC 2989  
2254 gaagacaaatccaaagcccgatgcggtgatatataataataacacagcttttaacctaaaa 2313  
Db 2990 AAAGACAAATPCCAAACGCTGCCAGCATTAATGATATTAATTAACAGCGCTTTAACTAAAA 3049  
2314 aataatagcaactccgttggcttctccactataaacactgttgactttatcgatggc 2373  
Db 3050 AATAAATAACAACCCATGACTTTGCTCCACTTATGACATTTGACTTTGGCAATGGC 3109  
2374 aatgccacacccagtaagtgacttacgtgaacaccaaataaacacacagtaagtaacttat 2433  
Db 3110 AATGCCACACCCAGTAAACCCATGATACCGCTTAACAAAAACCAAGTAAAGTGGTATAT 3169  
2434 gatgtcaattgtgatgagaaaaacattgaactcaacagggcgaataatggcgaagacacacaaa 2493

3170 GATGTGAATGTGGATGATACCAACCATTCATCTAACAGGCACTGATGACAA---TAAAAA 3226  
2494 attggcggtcaaaacaccacactgaccacacacacacacacacacacacacacacacacac 2547  
Db 3227 CTGTGGCTCAAAACCAACCAACTGACAAAACAAGTGTCTAATGTTATACAGCAACTAAC 3286  
2548 ----aaetttagtaccaccgataacgatgccccttgttlaagccaaaagacatcgccgaaaaat 2604  
Db 3287 TTTTAATCTTTAACTCTAGTGTGAAGATGCCCTTGTTTAACGCCAAAGACATCGCCGAAAAAT 3346  
2605 ctaaacacccctagcgaagaaattcacacacacacacacacacacacacacacacacacacac 2664  
Db 3347 CTAAACACCCCTAGCCAGGAAATTCACACCACCAAGGACACAGACACCGCCCTACAA 3406  
2665 acctttaaagtcaaaaagacgctgca-----actgtagtcgaacacacacacacacacacac 2718  
Db 3407 ACCTTTACCGTTAAAAAGGTAGATGAATAATAAATTAATGCTGATGACGCCAAGCNCATCAC 3466  
2719 aaagatggttac 2778  
Db 3467 GTGGGTCAAAAGAACGCAAAATAATCAAGTCAACACCCCTAACACTCAAAAGGTGAACACGGT 3526  
2779 ctacacggttgcaccataaagaatgtagcgtttacaccttggcattacacacacacacacacacac 2838  
Db 3527 CTTAATATTAAAAACCGACAAAAATGGTACGGTTACCTTTGGCATTAACACCAACGCGGT 3586  
2839 cttaaagcggcgacagcaccactctaaacacacacacacacacacacacacacacacacacacac 2898  
Db 3587 CTTAAGCGCGCAAAAGCAC---CTAAGCAGCGGTGGCTGTCTATTAANAACCCCACT 3643  
2899 agtaacgaacaaatcccaagtcggtgctgtagtggcgtagtggcgtagtggcgtagtggcgtagt 2955  
Db 3644 GGTAGCGCAACAAATCCTAAGTCTGCTGTGATGGCGTGAAGTTTGCCAAAGGTAAATAAAT 3703  
2956 ggttaattcaagcactgcgctgtagcgaacacacacacacacacacacacacacacacacacacac 3015  
Db 3704 GGTGTTGTAGTGTCTGGCATTTGATGGCAACCTCGCATTTACAGAGATGAATTTGGCTTT 3763  
3016 actggggcgaatggtcacttgatccac 3075  
Db 3764 ACTGGGACTAATGGCTCACCTTGATAAAAAGCAAAACCCCACTTAAGCAAAAGAC----- 3814  
3076 gtgggtgaagttgaattaccacacactggtcattacacacacacacacacacacacacacacacac 3135  
Db 3815 -----GGCATTAACGACAGGTGGTAAAAAGATTTACCAAC 3847  
3136 attcaatcaggtgatatattaccac 3195  
Db 3848 ATTCAATCAGGTGAGATTGCCAAACACAGCCATGATGCTGTGACAGCGCGCAAGATTAT 3907  
3196 gatttaaaacccgaacttgaagcaaaatcaacacacacacacacacacacacacacacacacacac 3255  
Db 3908 GATTTAAAAACCGAATTTGAAAAACAAATCAGCAGTACTGCCAAAAACAGCACAAAACTCA 3967  
3256 ttacacgaattctcagtagcagatgaac 3315  
Db 3968 TTACACCAATTTCTCAGTAGCAGATGAACAAGGTAAATAAATTTACGGTTAGTAAACCCCTTAC 4027  
3316 tccagttatgacacctcaagacacctctgctcatcaccttgcagctgaaacacacacacacacacac 3375  
Db 4028 TCCAGTTATGACACCTCAAAAGACCTCTGATGTCATCACCCTTTCAGGTGAAAACGGCATTT 4087  
3376 accaccaaggttaataaaggtgtggtgctgctggtgattgacacacacacacacacacacacacac 3435  
Db 4088 ACCACCAAGGTAAATAAAGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 4147  
3436 acgcctaaagctgacctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 3495  
Db 4148 ACGCCTAAGCTGACCGTGGGTAAATAAATGGAAGGCAATTTGTCATTGACAGCCCAAAAT 4207  
3496 ggtcaaaatcacatcacaggaactaagcaacacactctagctaatgtttaccaatga-----t 3549



QY 5707 gtaaacattccagacatcaaaaaagagaccccaaatctcaggttcatcatataaccgcactgtc 5766  
Db 5447 GTAAACATTCCGACATCAAAAGAGACCCAAATTCAGTTTCATCATCTAAACCGCACTGTC 5806  
QY 5767 atcaagcagcagcaggttacttgcgttaaggttaataacgataaccgaaacacttgcact 5826  
Db 5507 ATCAAGCAGCGACGGTACTTGGCGTAAAGGTAAATACGATACCGAAACACTTGGCACT 5866  
QY 5827 ggttggttacaagtggcggtggtataagacggaacgctaacgctaacgcttaagcaatggt 5886  
Db 5567 GGTGGTATACAAGTGGCGGTGGATAAAGACGCAACGCTAACGGCGATTTAAGCAATGTT 5626  
QY 5887 tgggtcaaaaccccaaaagatgacagcaaaaaagccctgctcaccactataacgcgcga 5946  
Db 5627 TGGGTCAAAACCCAAAAGATGCGCAAAAAGACCCCTGCTGCCACTTTATAACGCGCA 5686  
QY 5947 ggtcagaccactatttgaccacaaccccgagagccattgacagataaatgaacaa 6006  
Db 5687 GGTCAAGCAACTATTGTGACCAACACCCCGAGAGCCATTGACAGATAAATGACAA 5746  
QY 6007 ggtatccgcttcttccatgctcaacgattggcaatcaagagcctggttacaggcgctaac 6066  
Db 5747 GGTATCGCTTCTTCCATGTCACGATGGCAATCAAGAGCCTGTGGTACAAAGGCGTAAC 5806  
QY 6067 ggcattgactcaagtcctcagcgaagcactcagtgccgtaggtttccaggccaaggca 6126  
Db 5807 GGCATTGACTCAAGTGCCTCAGGCAAGCACTCAGTGGCGATAGTTTCCAGGCGCAAGCA 5866  
QY 6127 gatggtgaagcccggttgcattagcagacaaacccaagcagcgaacaaatccatgcc 6186  
Db 5867 GATGGTGAAGCCGCGTGGCCATAGGCGAGACAAACCCAGCAGCAACCAATCCATCGCC 5926  
QY 6187 atcgggtataacgcacaagccacgggcatcaatccatccatcgccatcggttacaggcaatgtg 6246  
Db 5927 ATCGGTATAACGCACAAAGCACCGGGCGATCAATCCATCGCCATCGGTACAGGCAATGTG 5986  
QY 6247 gtaacagtaagcactctggtgccatcgccagcccaagcactgttaaggctgataacagt 6306  
Db 5987 GTAGCAGGTAAAGCACTGTGTGGCATCGCGGACCCAAAGCACTGTAAAGGCTGATAACAGT 6046  
QY 6307 tacagtggtgtaataacaaccagtttatcgatgccactcagaccgatgtcttgggtg 6366  
Db 6047 TACAGTGTGGGTAAATACACACAGTTTACCGATGCCACTCAACCGATGTCTTGTGTG 6106  
QY 6367 ggcataacatcacctgacggaagtaactcgtgtgctttaggttcaaaactctgcccac 6426  
Db 6107 GGCAATAACATCACCGTGACCGAAAGTAACTCGCTTGCCTTAGTTTCAAACTCTGCCATC 6166  
QY 6427 agtcaggacacacagcaggacacaaagccaaaaatctgacggcacagcaggtacaacc 6486  
Db 6167 AGTCAGGCACACAGCAGGCACACAAAGCCAAAATACTCAGCGCACAGGATACAAAC 6226  
QY 6487 accacagcaggtgcaacagtgacggttaaggtcttgctggacaaacgcggttgggtgcg 6546  
Db 6227 ACCACACAGGTGCNACCGGTACGGTTAAAGGCTTTGCTGGACAAACGCGGTTGTTGCG 6286  
QY 6547 gttcccggtggtgctcaggtgtgaacgcggtatcaaaaatgtggcagcaggtgaggtc 6606  
Db 6287 GTCTCCGTGGTGCTCAGGTGTGAACCGCGTATCCAAAAATGTGGCAGCAGGTGAGGTC 6346  
QY 6607 agtcgcaacagcagcagtgccggtcaatggtagccaggttgataaaaagcccaagcatt 6666  
Db 6347 AGTCCCAACAGCAGCGGTGGTCAATGGTAGCCAGTTGTACAAAGCCACCCAAAGATT 6406  
QY 6667 gccaacgcaacaaatgactgacctatcgatccacaaacgaaataaagcgaatgca 6726  
Db 6407 GCCNACCAACCAATGAGCTTGACCATCGTATCCACCAAAACGNAATAGGCCAATGCA 6466  
QY 6727 gggatttcacagcagtgccggttccatgccaacagccctacattctctggcagatcc 6786  
Db 6467 GGGATTTTCATCAGCGATGGCGGTGCGCTCCATGCCAAGCCCTACATTCTCGGCAGATCC 6526  
QY 6787 atggttaccgggggtatttgcaccccaacagcgggtcaaggtgcggtgaggtggtgactgcg 6846

Db 6527 ATGTTTACCAGGGGTATGTCACCCACACAGGTCAAGGTGCGGTGGCAGTGGGACTGTC 6586  
QY 6847 aagctgcgataaattgctcaatgggtatttaaaatcaatggttcagcgatccccagggc 6906  
Db 6587 AAGCTGCGGATATGTCATGGGTATTTAAATCAATGTTTTCAGCCGATACCCAGGC 6646  
QY 6907 catgtagggcgagcagttggtcaggttttcaacttt 6942  
Db 6647 CATGTAGGGCGGCGAGTTGTGTGAGGTTTTCACTTT 6682  
  
RESULT 6  
AR040716 6973 bp DNA PAT 29-SEP-1999  
LOCUS Sequence 1 from patent US 5808024.  
DEFINITION AR040716  
ACCESSION AR040716  
VERSION AR040716.1 GI:5960079  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 6973)  
AUTHORS Sasaki, K., Harkness, R.E., Loosmore, S.M. and Klein, M.H.  
TITLE Nucleic acids encoding high molecular weight major outer membrane protein of moraxella  
JOURNAL Patent: US 5808024-A 1 15-SEP-1998;  
FEATURES Location/Qualifiers  
source 1..6973  
BASE COUNT 2265 a 1555 c 1533 g 1620 t  
ORIGIN  
  
Query Match 37.2%; Score 2580.6; DB 9; Length 6973;  
Best Local Similarity 66.9%; Pred. No. 0;  
Matches 4803; Conservative 0; Mismatches 1104; Indels 1270; Gaps 29;  
  
QY 1 atgaatcacatctataaagtcattctttaacaagccacagcgacatttatggcgtggca 60  
Db 542 ATGAATCACATCTATAAGTCATCTTTAAACAAAGCCACAGGCACATTTATGGCAGTGGCA 601  
QY 61 gagtgcgcaaaatccccacagc-ggagggagtagcagtagtaccgcaggaaggtgggcag 119  
Db 602 GAGTACGCCAAATCCACAGCACGCGGGGGGGTAGTGTGCTACAGGGCAAGTTGGCAG 661  
QY 120 ctctcctgtcatccgctgactgctgttgcacgctcgctatcctagtgcgtgcgac 179  
Db 662 TGTATGCACCTCTGAGCTTTGCCCGCTATTGCCGCGCTCGTCTCGTGTGCGGTGCAAC 721  
QY 180 gctcaatggcagtgcttatgtctcaaa-----ataatagcaagatcgattggt----- 228  
Db 722 GCTCAGTGGCAGTGTCTTATGCTCAAAAAAAGATACCAACATATCGCAATTGCTGAACA 781  
QY 229 ----accacaggcaacaatgacaatgctcggtgtagcaatgaaagcatccattggtatgg 284  
Db 782 AAACACGCCAAGAGCGCTCAGGCACTGGCAAGCGGAGCGGTGATCGAGCCATTGCTATTGG 841  
QY 285 tagctctgtaaggcacatcccaatcagctcattgctcattcgtggtgtagcaaacacagatcc 344  
Db 842 TGAANAATGCTTAACCAAGCGGCTCAAGCCATCGCCATCGGTAGTAGTAA---TAAAC 898  
QY 345 tcgtaatacagcggttaatacagaagcgaggttccccacgccaaggttaaagagtcacatccg 404  
Db 899 TGTCAATGGAAGCAGTTTGGATAAGATAGGTACCGATGCTACGGGTCAAGAGTCCATCGC 958  
QY 405 catcgggtggtgtagtactgctgaggtgtagtgcctcagttgcccattggtgtagtgactt 464  
Db 959 CATCGGTGGTATGTAAGGCTAGTGGTGTGATGCTCGCTCGATTGCCATCGGTAGTACTT 1018  
QY 465 atatttgataggaagtagcactaactctaaatatccaaatgg---tccttttagcactct 521  
Db 1019 ACATTTGCTTGATCAGCATGTTGTAATCTCTAAACATCCGAAAGGTACTCTGTATTACGATCT 1078

Qy	1518	tggcaacagacccttaaacacaggtggcttgactgtttaataacacacactgggtgtagcaaa	1577
Db	2156	TGCCAAAGAGGTAGCAGTGTCTACAGATGGCGTTTACCATCGAACAGCTCAAGCCGCCCAAGCC	2215
Qy	1578	taacaaatccaagtcggtgctgtagtggtcattaaattggcagatgtgaattgaatgtatc	1637
Db	2216	TACTTTAAACGACGGCGCTGCGATCAGTGTCCACCTACTGMAATATACAGTTGATGTATAA	2275
Qy	1638	aaatgcgcgcaaaatcgcgactactcgtattataccgaagagagaaatggcttgcgtgc	1697
Db	2276	GAGTGGCAATTTACCGGCCCAACTTCAACATTTGGCGTGAAACCCACCGAGCTTACAG	2335
Qy	1698	tgaaggtaaaagtgtataa-----	1715
Db	2336	TCATGGCACTAGTGATAAAATTTAGTGTAAAGGTAGTGTACGAACAATAGCTTTAGTTAC	2395
Qy	1716	-----aaagtcaccataatttgataaaaaaacaactcaagtggtgggtg	1760
Db	2396	CGCCGAACATTTGCCAGCTATCTAAATGAAGTCAATCGACGGCTGACAGTGTCTTACA	2455
Qy	1761	taaaattaccaaaagacagtcgattaaatgcaggtgatcaaaagatcagtaattgaaga	1820
Db	2456	RAGCTTTACCGTTAAAGAAGAAGACGATGATCAGCCCAACGCTATACCGTGGTTAAAGA	2515
Qy	1821	tgaacgacgataaccgatgcagtcactataaacagcttaaacagtcacaacaagcgc	1880
Db	2516	TACGACAAAAAATGCGCGCGCAGTCAGCATCTTTAAACTCAAAAGTTAAACCGGTCTAAC	2575
Qy	1881	gcagcgtgacctacaa-----agctctctctatcgtgatgaaaggtcagga	1928
Db	2576	GGTTGCTACCAAAAAAGATGGTACGGTTACCTTTGGGCTTAGCCAAAGATAGCGGTCTGAC	2635
Qy	1929	attacgattagtaactgtattcttaattgaatgaataaccacacatacctttgagaccatcac	1988
Db	2636	CATTGGCAAAAGACCCCTAAACAACGATGGCTTGACTGTTAAAGATACCAACGACAAAT	2695
Qy	1989	ctttgcaggtgaaacgagcatcagtcagcaatgacatagccaaaggtaaaagtcagaag-	2047
Db	2696	CCAAGTCGGTGTAAATGGCATTTAAATTTACTAATGTGAATGGTAGTAATCCAGGTAAGTGG	2755
Qy	2048	--ttggtattgacccaatacaatgggtctcacccagcctaagctgacggtgggtgagcgataa	2105
Db	2756	CATTGGCAATACCGTCGCGATTACCCAGAGATTAATTTGGCTTTGCTGGTCTGTATGGTGC	2815
Qy	2106	agatgtaaactcaattggttattgagcaag-----	2137
Db	2816	AGTTGATACAAACAAACCTTATCTTTGATCAAGACAAAGCTACAAAGTTGGCAATGTTAAGAT	2875
Qy	2138	-----tggtagcgttaacgacacccaacaaacatcattagagattgtccccaacactgccc	2192
Db	2876	TACCAACACTGGGCATTAACGAGGTGGTTAAAGCCATCACAGGGCTGTCCCAACACTGCC	2935
Qy	2193	tagcattacaaatgcaggtggcgtagcgcacacagaaacaggggcaatacaataccacgcga	2252
Db	2936	TAGCATTGGCGATCAAGTAGC---CGCAACATAGAACTGGGCAATACAAATCCA---AGA	2989
Qy	2253	cgaagcaaatcaaaagccgcagtagtcggtgataattaacacagcgttttaaccta	2312
Db	2990	CAAAAGCAAAATCCACGCTGCCAGCATTAATGATATATTAATACAGGCTTTAAACCTAAA	3049
Qy	2313	aaataatagcaactcogttggcttgtctccacttataacactgtttgactttatcgatgg	2372
Db	3050	AAATAATAAACACCCATTGACTTGTCTCCACTTATGACATTTGTTGACTTTTGGCAATGG	3109
Qy	2373	caatgccaccacgcgttaagttaacttacgtgaaaccaaatacaaacacagtaagaactta	2432
Db	3110	CAATGCCACCAACCGCCACAGTAAACCCATGATACCGCTAACAAACCCAGTAAGTGGTATA	3169
Qy	2433	tgaatgaattggtatgagaaacattgaactcaacagggcgaataatggcgaagcaacaaa	2492
Db	3170	TGATGTGAATGTGATGATCAACCATCTCATCTAATCAAGGACACTGATGACAA---TAAAAA	3266
Qy	2493	aattgcctcaaacaccaccactgagccacacacaaatgcttaaagtaaaacacacc-----	2547



QY 4686 caacacccataacacaaagtaactcaggtgctgcattaccccttctggttaacccgatgc 4745  
Db 5322 -----  
QY 4746 caatggaagcccatcaatggcaccgatggcaagcccccaaaaagccatcaagggcgccga 4805  
Db 5322 -----  
QY 4806 tggtaatactatcaagccaaagccaaagcgtaoctgtggcaaaagatggcaagcccat 4865  
Db 5322 -----  
QY 4866 caccgatgggacaaaacttgccaatctggcagctcatggcaaaccccttgatgcaggcca 4925  
Db 5322 -----  
QY 4926 tcaagtgggcaagcctaggcggaactcagatgccatcccttaacccaatcaatgc 4985  
Db 5322 -----  
QY 4986 cacttggccaaaatlgacacacacaaacacaggttaatgccaatgcaggccaaagcccaag 5045  
Db 5322 -----  
QY 5046 tctgccagcctatcagcagcacagcaaaagtaatgctgcaggtgtcaaaagatgtgtataa 5105  
Db 5322 -----  
QY 5106 tgtaggctttaacttgcagaccacacacaaatcacaaatcaagtggactttgtcaaaagcctatgatatac 5165  
Db 5322 -----  
QY 5166 cgtcaacttltgtaattggtacaggtgccgacatcaaaagcgtgcgtaglgctgatggcac 5225  
Db 5322 -----  
QY 5226 gatgataacatcaccgtcaacacccgcttagcagcgacgatgatggcaatgtgt 5285  
Db 5322 -----  
QY 5286 tatcaaaagcaaaatggttaagttctacaagcagacgacacctcatgccaaaagcgtcact 5345  
Db 5322 -----  
QY 5346 aaaagcaggcaaatcagccagtgatgccaaaactccaactggtcttaagccttlttaaccc 5405  
Db 5322 -----  
QY 5406 caatgctgtaaaagcagtagcagcgatgcagtggtctcttaataaacttatcaaaagcgt 5465  
Db 5322 -----  
QY 5466 atttaaaatccaaagatggtacaactactactaccacagtaagctctgatggcatccta 5525  
Db 5322 -----  
QY 5526 agggcaagataacagcagcatcaacctaaagcaaaagatgggctgaatgtaggcgtaaggt 5585  
Db 5322 -----  
QY 5586 cateagcaatggtgtaagcacaaaagacacagcagcgtgccatgttacaacagtttaaa 5645  
Db 5327 CATCAGTATGTGGGCAAAAGGCACAAAAGATACCGACGCTGCCAATGTACAAACAGTTAAA 5386  
QY 5646 cgaagtacgcaacttgttgggtcttggtaattgctggtaataatgataacgctgcagcaatca 5705  
Db 5387 CCAAGTACGCAACTTGTGGGTCTTGGTAAATCCTGGTAATGATAACGCTGACGGCAATCA 5446  
QY 5706 ggttaacattgccgacatacaaaaaagacccaaattcaggttcatcatctaaccgcactgt 5765  
Db 5447 GGTAAACATTTGCCGACATCAAAAAAGACCCAAATTCAGGTTTCATCTCACTAACCCGCACTGT 5506

QY 5766 catcaaaagcaggcacggtactcttgccggttaaggtataacgataccagaaaaaacttgcac 5825  
Db 5507 CATCAAAAGCAGGCACGGTACTTGGCGGTAAGGTAATAACGATACCGAAAAATTTGCCAC 5566  
QY 5826 tgggtgtataaagtggtgataaagacgcaacgctlaacggtcgatcttaagcaatgt 5885  
Db 5567 TGGTGGTATACAAGTGGCGTGGATAAAGACGGCAACGCTTAACGGCANTTTAAGCAATGT 5626  
QY 5886 ttgggtcaaaaccccaaaagatggcagcaaaaaagccctgctgcacttatcaacccgc 5945  
Db 5627 TTGGGTCAAAACCCAAAAGATGSCAGCAAAAAGCCCTGCTGCCACTTATAACGCCGC 5686  
QY 5946 aggtcagacaactatttgaccacaaccccgagagaagccattgacagaataaataaaca 6005  
Db 5687 AGGTGAGACCAACTATTGTACCAACAACCCCGCAGAGCCATTGACAGATAAATGAACA 5746  
QY 6006 aggtatccgctcttccatgtcaacgatggcaatcaagagcctgtgtcaaaagggcgtaa 6065  
Db 5747 AGGTATCCGCTTCTTCATGTCAACGATGGCAATCAAGAGCCTGTGTCAAAAGGCGCTAA 5806  
QY 6066 cggcattgaactcaagtgcctcaggcaagcactcaagtggcgatagggttccaggccaaagc 6125  
Db 5807 CGGCATTGACTCAAGTGCCTCAGGCAAGCACTCAAGTGGCGATAGGTTTCAGAGCCAAAGCC 5866  
QY 6126 agatgtgaagcccgcttgcctataggcagacaaaacccaaagcaggaacccaatccatcgc 6185  
Db 5867 AGATGGTGAAGCCGCCGTGTCCTAGTAGCAGACAAACCCAAAGCAGCAACCAATCCATCGC 5926  
QY 6186 catcgtgtataacacacaaagccacggcgatcaatcaatccatcgatcagcagcaatgt 6245  
Db 5927 CATCGGTGATAAGGCACAAAGCCAGCGGCGATCAATCCATCCCATCGGATACAGGCAATGT 5986  
QY 6246 ggttaacaggtgaagcaactctggtccatcgcgacccagcactgttaaggtgtataacag 6305  
Db 5987 GGTACGAGGTAAAGCACTCTGGTGCCATCGGCGACCCAAAGCACTGTTAAGGCTGTATACAG 6046  
QY 6306 ttacagttgggttaataacacacagctttatcgatgccactcagaccgatgtctttgggt 6365  
Db 6047 TTACAGTGTGGGTATATACACCACTTTACCGATGCCACTCAAAACCGATGTCTTTGGTGT 6106  
QY 6366 gggcaataacataccgtgacccgaaagttaactcgttcccttaggttcaaaactctgcat 6425  
Db 6107 GGGCAATAACATCACCGTGACCGAAAGTAACCTCGGTTGCCCTTAGGTTTCAAACTCTGCCAT 6166  
QY 6426 cagtcagcagcacacgcagcagcacacaaagcccaaaaaactcagcgcacagcaggttacaac 6485  
Db 6167 CAGTGCAGGCACACAGCGCAGGCACACAAAGCCAAAATCTGACGGCACAGCAGGTACAAAC 6226  
QY 6486 caccacagcaggtgcaaacaggtacggttaaaagcttctgctggacaaaagcgggttggtgc 6545  
Db 6227 CACCACAGCAGGTGCAACCGGTACCGTTTAAAGGCTTTGCTGGACAAACGGCGGTTGGTGC 6286  
QY 6546 ggtctcgtgggtgctcaggtgctgaacgcggtatccaaaaatgtggcagcaggttgaagt 6605  
Db 6287 GGTCTCCGTGGGTGCTCAGGTGCTGAACCGCGTATCCAAAATGTGGCAGCAGGTGAGGT 6346  
QY 6606 cagtcacacagcagcagatgctggttcaatggtagcagttgtacaaaagccacccaagcgt 6665  
Db 6347 CAGTGCACACAGCAGCAGGTGCGGTCAATGGTAGCCAGTGTGTACAAAGCCACCAAGCAT 6406  
QY 6666 tggcaacgcaaccaaataagcttgacctatccaccacaaaacgaaaaataaagcccaatgc 6725  
Db 6407 TGCCAACGCAACCAATAGAGCTTGACCATGTATCCACCAAAAGCAAAATAAGGCCANTGC 6466  
QY 6726 aggaatttcatcagcagatggcgatggctccatggcacaagcctacattccttggcagatc 6785  
Db 6467 AGGGATTCATCAGCATGGCATGGCGATGGCTCCATGCCACAAGCCTACATTCCTTGGCAGATC 6526  
QY 6786 catggttacccgggggtatttgcacccacacacaggttcaaggtgcggtgagtgaggactgc 6845  
Db 6527 CATGGTTACCGGGGTATTGCCACCCACACAGCGTCAAGGTGCGGTGGCAGTGGGACTGTC 6586  
QY 6846 gaagctgtcggataatgggttcaatgggtatttaaaatcaatgggttcagcgccgatacccaagg 6905



```
Db 6587 GAAGCTGTGGGATAAATGGTCAATGGGTATTTAAATCAATGGTTACGCCGATACCCAAAG 6646
QY 6906 ccaatgtagggcgaggttggtgaggttttcaacttt 6942
Db 6647 CCATGTAGGGCGGCAGTTGGTGCAGGTTTCACTTT 6683

RESULT 7
AX079913 6973 bp DNA PAT 22-FEB-2001
LOCUS Sequence 1 from Patent WO0107619.
DEFINITION AX079913
ACCESSION AX079913
VERSION AX079913.1 GI:13159436
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE
1 (bases 1 to 6973)
AUTHORS Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.
TITLE Recombinant high molecular weight major outer membrane protein of
moraxella
JOURNAL Patent: WO 0107619-A 1 01-FEB-2001;
CONAUGHT LABORATORIES LIMITED (CA)
FEATURES
Source
1..6973
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 2265 a 1555 c 1533 g 1620 t
ORIGIN

Query Match 37.2%; Score 2580.6; DB 10; Length 6973;
Best Local Similarity 66.9%; Pred. No. 0;
Matches 4803; Conservative 0; Mismatches 1104; Indels 1270; Gaps 29;

QY 1 atgaatcaatctataaagtcatctttaaacaagccacagggcacattatggcgtggca 60
Db 542 ATGAATCACATCTATAAAGTCATCTTTAAACAAGCCACAGGCACATTTATGGCAGTGGCA 601
QY 61 gagtgcacaaatccacagc-ggagggagtagcagtagtaccgcgagcacagtgagcag 119
Db 602 GAGTAGCCAAATCCACAGCACGCGGGGGGGTACTGTGCTACAGGCGCAAGTTGGCAG 661
QY 120 ctctctgtcactgcgctgactgtgttgccacgctcgtctatcctctagtgtgcagac 179
Db 662 TGTATGCACTCTGAGCTTTGCCCGTATTGCCGCTGCTGCTCTCGTATCGTGCACAC 721
QY 180 gctcaatggcagtgcttatgctcaaa-----ataatagcaagatcgcatgttgg- 228
Db 722 GCTCAGTGGCAGTGGCTATGCTCAAAAAAAGATACCAACATATCGCAATTTGGTGAACA 781
QY 229 ----accacagcacaatgacatccctcgctagcaatgaagacatccattgtattgg 284
Db 782 AACACGCAAGAGCTCAGGCACTGCCAAGCGGAGGTGATCGAGCCATTCATATTGG 841
QY 285 tagcttctgaagccacatcccaatcaagctattgtatctggtggtagcaaaaccagatcc 344
Db 842 TGAATAATGCTAACGCACAGGGCGGTCAAGCCATCGCCATCGGTAGTAGTAA---TAAAC 898
QY 345 tctgaatcaagcggttaatacagaagcgaggttcccacgcgaaggttaagagtcctatcgc 404
Db 899 TGTCAATGGAAGCAGTTTGGATGAAGATAGGTACCGATGCTACCGGTCACAGAGTCCATCGC 958
QY 405 catcgtggtgtagtactgaggtgtagtgcctcgtcagttgcatggttagtgaactt 464
Db 959 CATCGGTGGTGTGAAGGCTAGTGGTGAATGCTCGATGCTGCTGCTCGGTAGTATGACTT 1018
QY 465 atattggataggaatagcactaactctaaataatccaaatgg---tctcttagcactct 521
Db 1019 ACATTGCTTGATCAGCATGGTAACTCTAAACATCCGMAAGGTACTCTGATTAACGATCT 1078
```

```
QY 522 tattcaaaaccatacagtagtattacgccaataacagagactcaaatggttctca---gaaata 578
Db 1079 TATTACGGCCATGCGAGTATTTAAAGAAATACGAAGCTAAAGAGTAATGATGTAAATA 1138
QY 579 tagacacagacagagaagacacgccagtagtactgagtgaggagccatgcatatgcaaa 638
Db 1139 TAGACGCACAACCCGCAAGCGGACACGCCAGTACTGCTAGTGGGAGCCATGTCTATGCGACA 1198
QY 639 gggtcattttgccaacgcttttgggtacacggttcaacagctggaaggaacattcttccttggc 698
Db 1199 GGCTCATTTTTCACAGCCCTTTGGTACACGGCAACAGCTTAAAGTGCCTATTTCCTTGGC 1258
QY 699 agtaggtcttcacgcgccaagccgaaaaagatatatacaatcgctattggttctaagcaca 758
Db 1259 AGTGGGTCTTTGCCGCCACAGCCGAGGGCCAAATCTACAATCGCTATTGTTGTTCTGCTGAAC 1318
QY 759 agctatacaattatgagcactagccttgggtgcagatactcagtagtattgattacgg 818
Db 1319 ATCTAGTCTGTTGGAGCGGATAGCCCTTTGGTGCAGGTACTCGTCTCAGCTACAGGCGAG 1378
QY 819 tattgcctaggttatggttctcagatccttaataataataataataa---ataataa 875
Db 1379 TATTGCCCTTAGGTCAAGGTTCTGTGTCACCTCAGAGTGATAATAATTCCTAGACCGGCTA 1438
QY 876 agcctatgtaccagaaggttaattgggttcaaacataaaaatcgtctaaagccaccggcaatgg 935
Db 1439 TACACCAAAATACCCAGGCACCTAGACCCCAAGTTTCAAGCCACCAATAATACGAAGCGGG 1498
QY 936 ttattttccattggttagtagcactatcaagcgtataaatcatcaatgctcgttcaggtta 995
Db 1499 TCCACTTTCCATTGGTAGTAACCTCTATCAACGTTAAMTCAATCATCTCGGTGAGGTGT 1558
QY 996 tgaggataccgtagcggtcaatgtggcacagctaaaaagcgttgagaaatctggcta---a 1052
Db 1559 TAATAAAACCGATGCGGTCAATGTGGCACAGCTAGAAGCGGTGCTGAAGTGGCTAAGGA 1618
QY 1053 gcgtcaaatctatttaaggggtgatataacggtactggtctaaagaaaaaacatgggca 1112
Db 1619 CGCTAGAAATTTACTTTTCAGGGTGATGATAACAGTACTGACGTAAAAATAGTGTGGATAA 1678
QY 1113 gactttaaccattaaaggtggtgagaccccaagcggaacagctaacccgataataaatacat 1172
Db 1679 TACTTTAACTATTAAAGGTGGTG---CAGAGACCAACGCAATTACCCGATATATATATCGG 1735
QY 1173 tgggtggttaacagataataactggtctgaaagtttaaacttgcataaaacccctaaagcgg 1232
Db 1736 TGTGGTAAAGAGGCTGATAATAATAGTGTGCTGAAAGTTAAACTTGCCTAAACTTTAAACAA 1795
QY 1233 tcttgaacaagtttagcaccacaaacctaaccgcagcgaagaaagttacggtag---gtag 1289
Db 1796 TCTTACTGAGGTGAATACAACACTACATTTAAATGGCCACACACAGTTTAAAGGTAGTAG 1855
QY 1290 tggtaataacacccgtgagctacaaagcgtgtttaacctttacc---caacaac 1343
Db 1856 TAGTAGTACTACAGCTGAATATTGAGTGATAGTTTAACTTTTACCCAGCCCAATACAGG 1915
QY 1344 aaatgcaagcacagacacaaacccgtctatggtcactgtagggcttaagtttactgataatc 1403
Db 1916 CAGTCAAAAGCACAAGCAAAACCCTGCTATGCGTTAATGGGTGAAAGTTTACTAATAATGC 1975
QY 1404 taatacggcaacttga---gatactactgtataccacaaagataaaatggtttttagcaa 1460
Db 1976 AGAAACAACAAGCAGCAATCGGCACCTACTCGTATTACACAGAGATAAAATTTGGCTTGTGTCG 2035
QY 1461 taagctcgtgtagagttgataaaaacaccccttctctgataaagacagctaaag--- 1516
Db 2036 AGATGGTGTGTTGATGAAGAAACAAGCACCCTATTTTGGATATAAAACAACACTTAAAGTGGG 2095
QY 1517 -----t 1517
Db 2096 TAGTGTGTGCAATTACCATACACAATGCGCATTTGATGCGAGGTAATAAAAAAGATCAGTAATCT 2155
QY 1518 tggcaacagcacccctaacaacacggtggcttgactgttaataacacccattggtgtagcaa 1577
```







Db	5507	CATCAAGCAGCGACGGTACTTGGCGTTAAAGGTAATAACGATACCGAAAACTTGCAC	5566	
Qy	5826	tggtagtgataaagtgagcgtgagataaagacgcaacgctaacgagcgtatattaagcaatgt	5885	
Db	5567	TGGTGGTATACAAAGTGGCGGTGGATTAAGACGCGCAACGCTAACGGCGATTAAACCAATGT	5626	
Qy	5886	ttgggttcaaaaccccaaaagatggcagcaaaaagccctgctgccacttataacgcgcg	5945	
Db	5627	TTGGGTCAAAACCCAAAAGATGGCAGCAAAAAGGCCCTGCTGCCCACTTTATAACGCCGC	5986	
Qy	5946	aggtcagacaactatttgacaacaaccccgacgaacccatttcacagataaataagaaca	6005	
Db	5687	AGGTACAGACCAACTATTTTGACCAACAACCCCGCAGAACCCATTACAGCAATAAATGAACA	5746	
Qy	6006	aggtatccgctcttccatgtcaacgatggcaatcaagagcctgtgggtacaaagggcgtaa	6055	
Db	5747	AGGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCGTGTGTACAAAGGCGGTAA	5806	
Qy	6066	cggcatgactcaagtgcctcaggcaagcaactcagtgcgatagtttccaggccaaagc	6125	
Db	5807	CGCATTTGACTCAAGTGGCCTCAGGCAAGCACTCAGTGGCGATAGGTTTCCAGGCGAAGGC	5866	
Qy	6126	agatggtgaagcgcgcttgccatagcgagacaacccaagcaggcaacccaatccatcgc	6185	
Db	5867	AGATGGTGAAGCGCGCTTGGCATAGGCAGACAAACCCCAAGCAGGCAACCAATCCATCGC	5926	
Qy	6186	catcggtgataacgcacagcaagcagcgaggaatcaatccatcgcatcggtacaggaatgt	6245	
Db	5927	CATCGGTGATAAAGCCCAAGCCACGGCGCATCAATCCATCGGCCATCGGTACAGGCAATGT	5986	
Qy	6246	ggttaacaggtlaagcaactctggtgccatcgccgacccaagcaactgttaagcgtgataacag	6305	
Db	5987	GGTAGCAGGTAAAGCACTCTGTGTCCTCGCATCGCGACCCCAAGCACTGTTAAGCGTGTAAACAG	6046	
Qy	6306	ttacagtggtgggtaataacaacagctttatcgatgcactcagacagatgtcttgggtgt	6365	
Db	6047	TTACAGTGTGGTAAATAACAAACAGTTTACCGATGCGCATCAACCCGATGTCTTGGGTGT	6106	
Qy	6366	gggcaataaatacaccgtgaccgaagtaactcgttgctttagtttcaaaactctgccaat	6425	
Db	6107	GGGCAATAAATACATACCGTGACCCGAAAGTAATCTCGTTGGCTTAAACTCTGCCAT	6166	
Qy	6426	cagtcaggcaacacacgagcagcaacaaagcaaaataatcgacggcacagcaggtacac	6485	
Db	6167	CAGTGCAGGCACACACGACGACACAAAGCAAAAATCTGACGGCACAGCAGGTACAAC	6226	
Qy	6486	caccacagcagtgcaacaggtacggttaaaggctttgtggcaaacgcggttggtgc	6545	
Db	6227	CACCACACAGCTGCACCGGTACGGTTAAAGGCTTTGCTGGACAAACCGCGTTGGTGC	6286	
Qy	6546	ggtctccgtgggtgctcaggtgctgaacccggtatcccaaaatgtggcagcaggtgaggt	6605	
Db	6287	GGTCTCCGTGGTGCCTCAGGTGCTGAACGCGTATCCAAAATGTGGCAGCAGGTGAGGT	6346	
Qy	6606	cagtcgaacagcaccgatcggtcaatggttagcaggttgatacaaaagcccccagacat	6665	
Db	6347	CAGTGCCACACGACCGATGCGGTCAATGTGTAGCCAGTGTGTACAAAGCCCAACCAAT	6406	
Qy	6666	ttccaacgaacccaatgagctgaccatcgatcccaacaaacgaataaagccaatgc	6725	
Db	6407	TCGCCAACCAACCAATGAGTTGACCATCGTATCCAGCAAAACGAAAATAAGGCCAATGC	6466	
Qy	6726	agggatttcaatgaagcgaatggcgaatggcgtccatcccaagcctacatctcgtgcagatc	6785	
Db	6467	AGGGATTTTCAATCAGGCGATGGCGGTGCCGTCCATGCCACAACCCATACATTCTGGCAGATC	6526	
Qy	6786	catggttacccgggggtattggccaccacaacaggttcaaggtcggtgagctggactgtc	6845	
Db	6527	CATGGTTACCGGGGTATTGGCCACCCACACAGGTCAAGGTGCGGTGGCAGTGGGACTGTC	6586	
Qy	6846	gaagctgctcgataatggttcaatgggtatttaaaataaatggttcagccgatacccaagg	6905	

RESULT 8

AX079914 5976 bp DNA PAT 22-FEB-2001

LOCUS Sequence 2 from Patent WO0107619.

AX079914

ACCESSION AX079914.1 GI:13159437

VERSION

KEYWORDS Moraxella catarrhalis.

SOURCE Moraxella catarrhalis.

ORGANISM Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.

REFERENCE 1 (bases 1 to 5976)

AUTHORS Loomore, S.M., Sasaki, K., Yang, Y.P. and Klein, M.H.

TITLE Recombinant high molecular weight major outer membrane protein of moraxella

JOURNAL Patent: WO 0107619-A 2 01-FEB-2001;

CONNAUGHT LABORATORIES LIMITED (CA)

FEATURES

source

1..5976

Location/Qualifiers

/organism="Moraxella catarrhalis"

/db\_xref="taxon:480"

BASE COUNT 1972 a 1360 c 1355 g 1289 t

ORIGIN

Query Match 35.8%; Score 2485.8; DB 10; Length 5976;

Best Local Similarity 66.7%; Pred. No. 0;

Mismatches 1067; Indels 1269; Gaps 28;

Matches 4675; Conservative

Qy 166 gtgacgtgagcgtcgaatggcgagctgttatgtctcaaaataatagca-----agatc 219

Db 1 GTGATCGGTGCACGCTCAGTGGCGAGTGTATGCTCAAAAAGATACCAACATATC 60

Qy 220 gcatttgga-----ccacaggaacaataagcaactcctcgctgagcaatgaagca 270

Db 61 GCAATGTGTGAACAAACACCCAGCCAGCCCTCAGGCACTCGCCAAGCGGAGGTATCGA 120

Qy 271 tccattgtattggttagtctgttaaggcacatgccaatcaagctattgctatcggttgg 330

Db 121 GCCATTGTATTGGTGAATAATGCTAAACGCACAGGCGGTCAAGCCATCGCCATCGGTAGT 180

Qy 331 agcaaacagatcctcgttaatacagcggttaatacagaagcaggttccccacgcgaaggt 390

Db 181 AGTAA---TAAACACTGTCAATGGAAGCAGCTTTGGNTAAGATAGGTACCGATGCTACGGGT 237

Qy 391 aaagagtcctatcgccatcggttggtgatgtactggtgagggtagagcctcgtatgccaatt 450

Db 238 CAAGAGTCCATCGCCATCGGTGTGATGTAAGAGCGTGTGTTGATGCTCGATTTGCCATC 297

Qy 451 ggtagtgatgaatttatatttgatagggaatagcactaaactataataatccaaatgg---t 507

Db 298 GGTAGTGTAGTACTTACATTTGCTTGTATCAGCATGTGTAATCTTAAACATCCGGAAGGTACT 357

Qy 508 ctcttttagcactcttattcaaaacccatcacagtattacgcgaataacacgagactcgaatggt 567

Db 358 CTGATTAAACGATCTTATTAAACGCCCATCGAGTATTAAAGGAATAACGAAGCTCAAAAGAT 417

Qy 568 tctca---gaaatatagacgcagcagcagaagacacgcagccagctacgtgagagcc 624

Db 418 AATGATGTAATAATATAGCGCACACCCGGAAGCGGACACGCCAGTACTGCAGTGGGAGCC 477

Qy 625 atggcataacaaaggtcatctttgccaacgctttgggtacacggttcaacagctgaagac 684

Db 478 ATGTCTATATGCACAGGCTCATTTTTCACACGCTTTTGGTACACGGGCAACAGCTAAAGT 537

Qy 685 aactattccttggtgagtaggtcttaccgcgaagcgcaaaaggatatacataatcgctatt 744



QY	2704	accatcacgctgggtaaaagatgggtacacaaacggaagccgtcaacactctaaaaact	2763	3775	aattgggataataaaacacattgaagtgaacagtgataaaaaacttggcgctcaaaaccacc	3834
Db	2746	GCCACGCCATCACCGTGGGTCAAAAGAACGCAATAATCAAGTCAACACCCCTAACACTC	2805	3787	ATGTGGATGATACACACCATTTGAAGTAAA---GATAAAAACTTGGGTAAAAACCCACC	3843
QY	2764	aaagtgaaacacggtcttaacgggttgcctccaaataaagatggtaaggttaaccttggcatt	2823	3835	acactgaccaaaaaaagtgctaatggtaatgca---accaaatttagtgcgcgcgctgac	3891
Db	2806	AAAGGTGAAAACGGTCTTAATAATTAAACCCGACAAAATGTACGGTTTGGCATTT	2865	3844	ACATTTACCAGTACTGGCACAGGTGCTAATAAATTTGCCCTAAGCAATCAACCTACTGGC	3903
QY	2824	aaacacaaacggtctttaaagcggcgacagcaccactcttaaaacaaagatggcttctct	2883	3892	gatgcccttgttaaagccagtgatcgcacccacttaataacaccttggctggcgacatc	3951
Db	2866	ACACACCAAGCGGTCTTAAAGCCCGCAAAAGCAC---CTAAACGAGCGGTGTCT	2922	3904	GATGCCCTGTCAAGCCAGTGTATCTGTCTATCAACACCTTATCTGGCGACATC	3963
QY	2884	attaaaaaccccgctagtaacgaacaaatccaagtcgctgctgctgctgctgctgctgct	2943	3952	caaacgcgaaggggcaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc	4011
Db	2923	ATTAAAAACCCACTGGTAGGCAACAAATCCAAAGTCGCTGTGATGGCGTGAAGTTGGC	2982	3964	CAAACTGCCAAAGGGCAAGCCAAAGCAACTGCGCCCAAGCCCAAGCCCAAGCCCAAG	4023
QY	2944	aaagt---tgataaggttaattcaagcactgctgctgctgctgctgctgctgctgctgct	3000	4012	aaacaggtctatctatcacagtcaccgataaagaagtaactcaagtcgaatgacaaggtcaa	4071
Db	2983	AAGGTTAATAATAATGGTCTTGTAGTGTGTCATTTGATGATGATGATGATGATGATGAT	3042	4024	AAATAAGTCTATCTATGACAGTACCGATACCAAGTACTATCAAGCCCAAAATGATGSCACA	4083
QY	3001	gatcaaatggcttactgggctaatggctcaacttgatcaccacaaacccacacacac	3060	4072	gtggacaaaaacaaagaagttgccaagaacaaactggctgcgccaagcccaacccccagat	4131
Db	3043	GATGAAATGGCTTTACTGGGACTAATGGCTCAGTTGATATAAAGCAAAACCCACCTAAGC	3102	4084	GTGTGATAAAACCAAGAAAGTTGCCAAAGACAAACTGGTGCGCCCAAGCCCAAGCCGAT	4143
QY	3061	aaagcaacgttaaagtggtgagttgaattaccaaactggcattaaacgaggtggt	3120	4132	ggcacattgctgctcaaatgaatgcaaatcagtcatttaacaaagaagcaagtaaatgatgcc	4191
Db	3103	AAAGC-----GGCATTAACGACGCTGGT	3126	4144	GGCAGATTGGCTCAATGAATGTCAATCAATCACTATCAATCAAGCAAAATGATGATGCC	4203
QY	3121	aaaaagattaccacattcaatcaggtgatattaccacaaacagcaatgatctgtgaca	3180	4192	aataaaagcaagcgatcaatgaagacaacgccttatcaaaagggttgaacacgcgcgc	4251
Db	3127	AAAAGATTACCAACATTCATCAGGTGAGATTGGCCAAAGACAGCCATGATGCTGTGACA	3186	4204	AAATAAAGCAAGGCATCAATGAAGACAAAGCCCTTTGTTAAAGGACTTGAAGAAGCCGT	4263
QY	3181	ggcggtgggttatgatattaaacacgacttgaaagcaaaatcaacagtgctgctaaa	3240	4252	aaagacacaaaaacaaaaacgcgcgagtaactgtggtgatitaaatgcgttgcgcaa	4311
Db	3187	GGCGGCAAGATTATGATTATAAACCAGAACCTTTGAAACCAAAATCAGCAGTACTGCCAAA	3246	4264	TCTGATAACAAACCAAAACCCCGCAGTAACTGTGGGTGATTATATGCGTTGCCCAA	4323
QY	3241	acagcaaaaaactcaatcacgaattctcagtagcagatgaacaaaggttaacacttaacg	3300	4312	acacgcgtgaccttggcaggggatacacagggcacaacgcgttaaaaaactggcgagactttg	4371
Db	3247	ACAGCACAAAACATTTACACGAATTTCTAGTAGCAGATGAACAGGTAAATAACTTTACG	3306	4324	ACACGCTGACCTTTGCAAGGGATACAGGCACAAACGGCTAAACAACTGGCGGAGACTTG	4383
QY	3301	gttagtaaccttactccagttatgacacctcaagacctctgctcaccctttgca	3360	4372	accataaagtgggcaaacagacacaaataagctaacccgataataacatcgggtggtga	4431
Db	3307	GTTAGTAACCCCTTACTCCAGTTATGACACCTTCAAGACCTCTGATGTCATCATTGGCA	3366	4384	ACCATCAAAAGGTGGCAACACAGACACCAANTAAAGCTAACCGATATAACATCGGTGTGTA	4443
QY	3361	ggtgaaacggcattaccacaaagtaataaaaggtgtgtgtgtgtgtgtgtgtgtgtgtgt	3420	4432	gcagtgactgagtgcttcaactgtcaaacctgcgaagacccaactcttaacagagctt	4491
Db	3367	GGTGAAACGGCATTTACCACCAAGGTAAATAAAGGTGTGGTGTGGCATTTGACCAA	3426	4444	GCAGTACTGATGGCTTCACTGTCAAACTTGCCAAAGACCTTAACCAATCTTTACAGCGTT	4503
QY	3421	accaaagcttaacacgcctgaactgagctgggtggtggtggtggtggtggtggtggtggt	3480	4492	aatgcaggtggcacacagaaattgatgaaagagcactctctttttagacgcgaacgggtcaa	4551
Db	3427	ACCAAAGGCTTAACACAGCCCTAAGCTGACCGTGGGTGATTAATTAATGCAAAAGGCATGTC	3486	4504	AATGCAGGTGGCACCATAATTTGATGCAAAAGCGGTGCTTTTGTAGACTCAAGCGGTCAA	4563
QY	3481	attgacagtaagatgggtcaaaaaccatcacaggaactgaagcaactctagctaatggtt	3540	4552	gcaaaagcaaacacccctgtgctaagtgccaatgggctgggaacctgggtggcaaacgcac	4611
Db	3487	ATTGACACCCAAATGGTCAAAATACCATCACAGGACTAAGCAACACTCTAGCTAATGTT	3546	4564	GCCAAAGCAAAACACCCCTGTGCTAAGTGCCATGGCTGGACCTGGGTGGC-----	4614
QY	3541	accaatga-----tgggtgcggacacgcactaaagcgaagggcttgcgaatgacacgcac	3594	4612	agtaacatcgggtgcagctgtgtgatgataacgagtgcggtgaacttttaagcagtttaata	4671
Db	3547	ACCAATGATAAAGGTAGCGTACGCCACACAGAACAGGGCAATATAATCAAGACGAAGAC	3606	4615	-----	4614
QY	3595	aaaacccgtgcgcagcagatgggtgatgtgctaaacgcagggctttaaacttgcagggcaat	3654	4672	gttgccaaaacggttcaacaaacctaaacacaaagtaactcaggtgctgtcattacccttt	4731
Db	3607	AAAACCCGTGCGCCAGCATTTGTTGATGCTAAGCGAGGCTTTAACTTGCAAGGCAAT	3666	4615	-----	4614
QY	3655	ggtgaaggggttgacttcttctccacttatgacactgttgactttatcgatggcaatgcc	3714	4732	gtggttaaccgatgccaattggcaagccccatcaatggcaccgatggcaagcccccaaaagcc	4791
Db	3667	GGTGAAGCGGTTGACTTTTGTCTCCACTTATGACACCGCTCAACTTGGCGATGGCAATGCC	3726	4615	-----	4614
QY	3715	accacgcctaaggtgaccttatgatgacacaaagcaaaacagtaaaagtggtctatgatgc	3774	4792	atcaagggcgcgatggttaatactatcacgcgaacgcgaacggcgctacacctgtgacaaa	4851
Db	3727	ACCACCCCTAAGGTGACCTATGATGACACAAAGCAACAGTAAAGTGGTCTATGATGATC	3786	4615	-----	4614
				4852	gatggcaagcccatcccgatgcggacaaacttgcgaactctgtggcagctcatgtgcaaaccc	4911











|||||  
Db 3965 AAACGCCAAGGGCAGCCAGCCAGCACTACAGAGGCTATGTGATGCTGATGSCA 4024  
Qy 4013 acaaggtcactatgacagttaccgataagagtagtactatcaagtcaatgacaaaggtgcaag 4072  
Db 4025 ATAAAGTCTATCTATGACAGTACCGATTAACAAGTACTATCAAGCCAAAATATGGCAGAG 4084  
Qy 4073 tggacaaaaaagaagttgccaagagcaaaactgtgcgccaagcccaaaacccagatg 4132  
Db 4085 TTGATAAAACCAAGTTGCCAAGACAAACTGGTCGCCCAAGCCCAACCCAGATG 4144  
Qy 4133 gcacattggctcaaatgaatgtcaatcagttcaatcaaaagagcaagttaaatgatgcca 4192  
Db 4145 GCACATTGCTCAATGATGTCAATCAGTCAATTAACAAGAACAAAGTAAATGATGCCA 4204  
Qy 4193 ataaaagcaaggtcatcaatgaagcaacgcctttatcaaaaggttgaacacgcgcca 4252  
Db 4205 ATAAAAGCAAGGCATCAATGAAGACACAGCGCTTTGTTAAAGGACTTGAAAAGCGCTT 4264  
Qy 4253 aagacacaaaacaaaacgcgcagtaactgtgggtgatttaagtgcgttgcgcaaa 4312  
Db 4265 CTGATTAACAAAACCAAAACCGCGCAGTAACCTGTGGGTGATTTAAATGCCGTTGCCAAA 4324  
Qy 4313 caccgtgacctttgcagggggtacagggcacacacgcgttaaaaaactggcgagactttga 4372  
Db 4325 CACCGTGCACCTTTGCAGGGGATACAGGCACAACGGCTAAAAAGCTTGGCGGAGACTTTGA 4384  
Qy 4373 ccaataaaggtgggcaaacagacacacaaataagctaacgataataacatcggtgtgtag 4432  
Db 4385 CCATCAAAAGGTGGGCACAAAGACACACCAATTAAGCTAACCGGATTAATCACTCGGTGTGGTAG 4444  
Qy 4433 caggtactgaggttcactgtcaaaacttgcaaaagacctaaccatcttaacagcgtaa 4492  
Db 4445 CAGGTACTGATGCTTCACTGTCAAACTTGCCTAAAGACCTTAACCAATCTTAACAGCGTTA 4504  
Qy 4493 atgcaggtggcaccagaattgataaaaaaggatctctctttttagacgcaaacggtgcaag 4552  
Db 4505 ATCAGGTGGCACCAAAATGTATGACAAAAGGGGTGCTTTTGTAGACTCAAGCGGTCAAG 4564  
Qy 4553 ccaaaagcaaacacccctgtgctaagtgccaatggcgctggacctgggtgcaaacgcataca 4612  
Db 4565 CCAAGCAACACACCCCTGTGCTAAGTGGCCATGCGGTGGACCTGGGTGGC----- 4614  
Qy 4613 gtaacatcgggtgcagctgttgatgataacgatgcggtgaacttgaagcagtttaataagaag 4672  
Db 4615 ----- 4614  
Qy 4673 ttgccaacacggTcaacaaaccttaacacaaagtaactcaggtgcgtcatattaccctttg 4732  
Db 4615 ----- 4614  
Qy 4733 ttgtaaccgatgcgaatggcaagcccatcaatggcaccgatggcaagcccccaaaagcca 4792  
Db 4615 ----- 4614  
Qy 4793 tcaagggcgccgatggtataataactatcacgccaacgccaacggcgtacctgtggacaaag 4852  
Db 4615 ----- 4614  
Qy 4853 atggcaagcccaTcaacgatgcggacaaaacttgccaatctggcagctcatggcaaacccc 4912  
Db 4615 ----- 4614  
Qy 4913 ttgatgcaggtcatcaagtgtggcaagccctagtcggcaactcagatgcacatcacccctaa 4972  
Db 4615 ----- 4614  
Qy 4973 ccaacatcaagtccaactttgccacaaattgacacacaaacacaggttaattgccaatgcaag 5032  
Db 4615 ----- 4614  
Qy 5033 ggcaagcccaagTctgcccagcctaTcagcagcacagcaaaagtaaatgctgcagtgtoa 5092  
Db 5147 TCCAGGCCAAGGCAGATGATGTTGAAGCCCGCGCTTGCCTATAGGCACACACCCANGCAGGCA 5206  
Db 4615 ----- 4614  
Qy 5093 aagatgtgtaaaatgtaggctttaacttgagaccatcaaatcaaatggaacttgta 5152  
Db 4615 ----- 4614  
Qy 5153 aagcctatgataccgtcaactttgtcaatggtacaggtgcccagatcacaagcgtgcgta 5212  
Db 4615 ----- 4614  
Qy 5213 gtgctgatggcaagatgagtaacatcacccgtcaacacgcgcttagcagcgaccgatgatg 5272  
Db 4615 ----- 4614  
Qy 5273 atggcaatgtgcttatcaaaagcaaaagatggttaagtctctacaagcagacgcctcatgc 5332  
Db 4615 ----- 4614  
Qy 5333 caaacggctcactaaaagcaggcaaaatcagccagtgatgccccaaactccaaactggcttaa 5392  
Db 4615 ----- 4614  
Qy 5393 gccttgttaaccccaatgctgttaaaagcagtcacagggcagtgctcttctaataact 5452  
Db 4615 ----- 4614  
Qy 5453 tatcaaaagcgtgatttaaatccaaagatggtacaacttactaccacagtaagctctgatg 5512  
Db 4615 ----- 4614  
Qy 5513 gcatacagtatccaagggcaaaagataaacacagcagcatcacccctaagcaaaagatgggtgaaatg 5572  
Db 4615 ----- 4614  
Qy 5573 taggcggtaaagtcacagcaatgtgggtaaaaggcacaaaaagacaccgacgctgccaatg 5632  
Db 4615 ----- 4614  
Qy 5633 tacaacagttaaacgaagtacgcgaactgtgtgggtccttggtaactgtgtaataagtaacg 5692  
Db 4667 TACAACAGTTAAACGAAGTACGCAACTTTGTTGGGTCTTTGGTAACTGCTGTAATATAACG 4726  
Qy 5693 ctgacggcaatcaggtaaaacattgcgcgacatcaaaaaagaccacaaattcagattcatcat 5752  
Db 4727 CTGACGGCAATCAGGTAAACATTTGCCGACATCAAAAAGACCCCAATTCAGGTTTCATCAT 4786  
Qy 5753 cttaacgcgactgtcatcaaaagcaggcacggttacttggcggttaaaaggttaataacgatacgg 5812  
Db 4787 CTAAACGGCAGTCTCATCAAAAGCAGGCAGCGTACTTTGGCGGTAAAGGTAATAACGATACCG 4846  
Qy 5813 aaaaacttgccaactgtggtgtacaagtggtgtgataaaagcagcgaacgcaacgctgacg 5872  
Db 4847 AAAAACTTGCCCACTGGTGTATACAAGTGGGCTGTGATTAAGACCGCAACGGTACAGCGG 4906  
Qy 5873 atttaagcaatgtttgggtcaaaaccccaaaaagatggcagcaaaaaagccctgctcgcca 5932  
Db 4907 ATTTAAGCAATGTTTGGGTCAAAACCCCAAAAAGATGGCAGCAAAAAGCCCTGCTCGCCA 4966  
Qy 5933 cttaaacgcgcaggttcagacccaactatttgacccaacccccccagagcattgaca 5992  
Db 4967 CTATTAACGCCGCGAGTCAACCAACTATTTTGAACCAACAAACCCCGCAGAAAGCATTTGACA 5026  
Qy 5993 gaataaatgaacaaaggtatccgcttcttccatgtcaacgatggcaatcaagagcctgtg 6052  
Db 5027 GAATAAATGAACAAGGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCCCTGTGG 5086  
Qy 6053 tacaagggcgtaacggcatttgactcaagtgcctcagggcaagcactcagtggtgcgataagtt 6112  
Db 5087 TACAAGGGCTTAACGSCATTTGACTCAAGTGCCTCAGGCAAGCACTCAGTGGGGCAGTAGGTT 5146  
Qy 6113 tccaggccaagcagatggtgaagcccgcttgcatagcagacaaaaccccaagcagga 6172  
Db 5147 TCCAGGCCAAGGCAGATGATGTTGAAGCCCGCGCTTGCCTATAGGCACACACCCANGCAGGCA 5206





```
Db 2101 AGCAAAAGCCCTGCTCGCACTTATAACGCCGAGGTGAGACCAACTATTGACCAAC 2160
QY 5971 aacccgcagaagccattgacagaataaataaacaaggtatccgctcttccatgtcaac 6030
Db 2161 AACCCGCGAAGCCATTGACAGNATAATGACAAAGGTATCCGCTCTTCATGTCAC 2220
QY 6031 gatggcaatcaagagcgtggtggtacaaagggcgtaacggcattgactcaagtgcctcaggc 6090
Db 2221 GATGGCAATCAAGAGCCTGTGTGTACAAAGGGCGTAACGGCATTTGACTCAAGTGCCTCAGGC 2280
QY 6091 aacactcaatgcataggtttccagggccaaagcgagatggtgaagccgctgtgccata 6150
Db 2281 AAGCACTCAATGCGATAGGTGTCAGGCCAAGCGAGATGTTGAAGCCGCGGTGTCATA 2340
QY 6151 ggcagacaaacccaagcaggcaaccaatccatcgccatcggtgataacgcacaaagccagc 6210
Db 2341 GGCAGACAAACCCAGCAGCAACCAATCCATGCCATCGGTGATACGGCAGCAGCCACG 2400
QY 6211 ggcgataatccatcgccatcggtacagggcaatgtggtacaggtgaagcactctgtgcc 6270
Db 2401 GGGCATCAATCCATCGCATCGGTACAGGCAATGTGTAGCAGTAAGCAGTCTGTGTGCC 2460
QY 6271 atcggcgaacccaagcactgttaaggctgataaacagtttaacagttggtgtaatacaaacag 6330
Db 2461 ATCGGCGACCAAGCACTGTTAAGGCTGATTAACAGTTTACAGTTGAGTGGTATAAACAACAG 2520
QY 6331 ttatcgatcgactcagacccgctgtcttttgggtgggcaataacatcacctgacccgaa 6390
Db 2521 TTACCAGTCCCACTCAACCCGATGCTTTTGGTGTGGCAATACATCACCGTGCACCGAA 2580
QY 6391 agtaactcggtgtccttagatgtaactctgccaactcagtgtaggcacacacgcagggcaca 6450
Db 2581 AGTAACTCGGTGTGCTTAGTTCAAACTCTGCCATCGTCAAGCAGCACACACGAGGCACA 2640
QY 6451 caagccaaaatactgacgcacagcaggtacaaccaaccaacagcaggtgcaacaggtacg 6510
Db 2641 CAGCCAAAATACTGACGCGACAGCAGGTACAAACCACCAACAGCAGCAGGTGCAACCGTACG 2700
QY 6511 gtaaaagcgttctggtgacaaaagcggttgggtggtgctccgtggtgctcaggtgct 6570
Db 2701 GTTAAAGGCTTTGCTGGACAAACGGCGGTTGGTGGGTCTCGGTGGTGGCTCAGTGTCT 2760
QY 6571 gaacgcgtatccaaaatgtygcagcaggtgagtgtagtgccaccagcaccgtagcggtc 6630
Db 2761 GAACGCGTATCCAAAATGTGGCAGCAGGTGAGTCACTGCTCAGTCCACGACCGATGCGGTC 2820
QY 6631 aatggtagcagttgtacaagccaccacccaaagcagcatgccaacgcaaccaatgagcttgac 6690
Db 2821 AATGGTAGCCAGTTGTACAAGCCACCCAAAGCATTGCCCAACGCAACCAATGAGCTTGAC 2880
QY 6691 catcgtatccaccacaaaataaagcccaatgacggggttttcacagcgatggcgatg 6750
Db 2881 CATCGTATCCACCAAAACGAAATAGGCCAATGCGAGGATTTCAATCAGCGATGGCGATG 2940
QY 6751 gcttccatgcacaagcctacattcctggcagatcccatggttaccgggggtattgcacc 6810
Db 2941 CGGTCCATGCGCACAAAGCCTACATTCCTGGCAGATCCATGTTTACCGGGGGTATTGCGACC 3000
QY 6811 cacaaggtcaaggtgcggtggcagtgagactgtcgaagctgtcgagataatggtcaatgg 6870
Db 3001 CACACGGTCAAGTGGCGGTGGAGTGGGACTGTCGAAGCTGTCGAAGTATGGTCAATGG 3060
QY 6871 gtatttaaatcaatggttcagccgaaccacaaagggcaggtgagggcgaggttgggtgca 6930
Db 3061 GTATTTAAATCAATGTTTCAGCGCATACCCAGGCCATGTAGGGGGCGCAGTTGGTGCA 3120
QY 6931 ggtttcaacttt 6942
Db 3121 GGTTTTCACTTT 3132
```

RESULT 11

```
AX079959
LOCUS 2448 bp DNA PAT 22-FEB-2001
DEFINITION Sequence 47 from Patent WO0107619.
ACCESSION AX079959
VERSION AX079959.1 GI:13159466
KEYWORDS Moraxella catarrhalis.
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 2448)
AUTHORS Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.
TITLE Recombinant high molecular weight major outer Membrane protein of moraxella
JOURNAL Patent: WO 0107619-A 47 01-FEB-2001;
CONNAUGHT LABORATORIES LIMITED (CA)
FEATURES
source 1..2448
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 826 a 569 c 556 g 497 t
ORIGIN
Query Match 20.5%; Score 1421.8; DB 10; Length 2448;
Best Local Similarity 87.2%; Pred. No. 0;
Matches 1648; Conservative 0; Mismatches 187; Indels 54; Gaps 6;
QY 2803 ggtacggttacccttgccattaacaccacaaagcggtcttaaaagcggcgacagcaccact 2862
Db 1 GGTACGGTTCCTTTGGCATTACACCAAGCGGTCTTAAAGCCGCGCAAGACCC--- 57
QY 2863 ctaaaacgaagtgcttctctattaaaaccccgctctagtaacgacaaatccaaagtcggt 2922
Db 58 CTAACCGACGGTGGCTGCTCTATTAAACCCCACTGGTAGCGAACAATCCAAAGTCGGT 117
QY 2923 gctgattggcgtgaagtttgccaaggt---tgataagggtaattcaagcactggcattgat 2979
Db 118 GCTGATGCGCTGAAGTTTGCCAGGTTAATAATAATGCTGTGTAGGTGCTTGGCATTTGAT 177
QY 2980 ggcacagcgcgtatcaccaagatcaaatggcttcttactggggcgctaatggcctactgat 3039
Db 178 GGCACAACCTCGCATACCAAGATGAAATTGGCTTTTACTGGGACTAATGGCTCACTTGAT 237
QY 3040 accaccaaaacccaccccttaaccaagacaaagcttaaaagtggggtgaagtgaataaccac 3099
Db 238 AAAAGCAAAACCCCACTTAAGCAAAAGAC----- 264
QY 3100 actggcattacgcaggtggtgtaaaagattaccaacattccaatcaggltgatattacccaa 3159
Db 265 ---GGCATTTAACGCAGGTGGTAAAAAGATTACCAACATTTCAATCAGGTGAGATTGCCAA 321
QY 3160 aacagcaatgatgctgacagcggtcggttttatgatttttaaaacccgaactgaaagc 3219
Db 322 AACAGCATGATGCTGTGACAGCGCGCAAGATTATGATTTTAAACCCGAACTTGAAAC 381
QY 3220 aaaaatcaagctgctgtaaaacagcaaaaactcattacacgaattctcagtagcagat 3279
Db 382 AAAATCAGCAGTACTGCCAAAACAGCACAACCTCATTACAGAAATCTCTAGTAGCAGAT 441
QY 3280 gaacaaggttaateacttaccggttagtaaaccttactccagttatgacacctcaaaagacc 3339
Db 442 GAACAAGGTAAATAATTTTACGTTTAGTAAACCTTTACTCCAGTTATGACACCTCAAGACC 501
QY 3340 tctgattcatcaacttgcaggtgaaagcgattaccaccaaggttaataaagggtgtg 3399
Db 502 TCTGATGTCAATCACTTTGACAGGTGAAACGGCATTTACCACCAAGGTAATAAAGGTGTG 561
QY 3400 gtgcgtgtgggcatttgaccacaaacaaagcgttaaccacgccttaagctgagcgtgggtaat 3459
Db 562 GTGCGTGTGGCATTGACCAAAACAAAGGCTTAACACGCGCTAAGCTGACCTGGGTAAAT 621
QY 3460 aataatggcaagcaggttgcattgacagtaaaagtggtcaaaataccatcacaggacta 3519
```

Thu Sep 13 14:17:47 2001

```

Db 1699 TTTGTAGACTCAAGCGTCAAGCCAAAGCAACACCCCTGCTCTAAAGTGCCAAATGGCGCTG 1758
QY 4591 gacctgggtggcaaacgcatcaatcaatcagtcgagctgttgatgataacgagtcgcggtg 4650
Db 1759 GACCTGGGTGGCAAGTGCATCAGTAATGTGGCAAGGACACAAAGATATACCGACGCTGCC 1818
QY 4651 aactttaagcagcttaataagtgagtgccaa 4679
Db 1819 AATGTACACAGCTTAACGAGTACGCAA 1847

RESULT 12
PMU277635 3900 bp DNA BCT 21-OCT-2000
LOCUS Pasteurella multocida mapA gene.
DEFINITION AJ277635
ACCESSION AJ277635.1 GI:10945096
VERSION MapA protein.
KEYWORDS Pasteurella multocida.
SOURCE Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
ORGANISM Pasteurella.
REFERENCE 1 (bases 1 to 3900)
AUTHORS Henderson, I.R., Nataro, J.P., Cappello, R. and Stein, C.
TITLE Evolutionary origins of the autotransporter proteins
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3900)
AUTHORS Henderson, I.R.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2000) Henderson I.R., Center for Vaccine
Development, University of Maryland School of Medicine, 685 W.
Baltimore St, MD 21202, USA
FEATURES
source
1..3900
/organism="Pasteurella multocida"
/strain="PM70"
/db_xref="taxon:747"
1..3900
/gene="mapA"
1..3900
/gene="mapA"
/function="putative autotransporter protein"
/codon_start=1
/transl_table=11
/product="MapA protein"
/protein_id="CAC14202.1"
/db_xref="GI:10945097"
/translation="MNKIYRTLMNAATQSWVYVSELAKAGKSAKSLVNSVGF
FTLASVVLGSGOVNAEVTGTVSGDDKYCFYNASSQSVICGDATTKTTDKDNK
PAKSVVIGFATNDETINVAIGAKSKSAIAIGDNALDNOALAIQONATNSD
WDISIGRQAGAEQTEVSAEGRNIAIGDALKRGKGVNNIALGTSAGDLAGTHVL
CTYVNADEAVSALTAGSTTKKEINAKETDNKYIEASNVVALGTRALATQLAAV
QQAFAFGQSVAVGNTKASQTAIATGSHAHATGSSIAIGTVNGNDRTARTLASD
FTIAMLGSAQATKSDAIACVNRNAAGINTVSYGNAGVTKTADAQSDPNKLEPAT
DAVFTGNKAGYKSNRMQVSLGDSGVGVGTENVITGNSAGKNTKGNNTVSLKES
QVNRHDPFALILEAGONIKGSDNIAIGKHAGRSADPTKLNINNTISLCKESVLK
NFGTAGNKAKTDGLASTAIGRNAENVGETNANTIAIGDSASADAGVILCTKAQAK
LTVDGKYGAYSIVGTEARAKIAQAAPAGKNENPKDAIAGTKAEAHYASTIALG
AKSDTKAQAYSIGVSNYNAKQYQAFGEAKTQNAATGSSIAFGTKAQTASASIAIG
GAETGFDGQALDSDSVALGREAKRQNALAFYKAVADHKDAVALGAGCAATAAE
GTNEATVNEFYSGPAGIKPATVSVGKKDAERTITNVAGRIDKTDSTDAINGSLYL
ALNALGVNLTIVTLVNLGDAALVKEGDEAGTLTMSNIGCTGCKTIDHAIIPAVNTAK
ASKTVPKSGDNITVTEEAADGSRITYTATKDKVDFSDVAVAGTFAQGEVKNIIIGDIT
YNPTEGEANTNIGTGASTIDEAKVNTTAKAKTEVVOGENIVTSAFGANGTV
YVATKAEVTFDKTITVSVVTDKNTNITLSNKTIGDNGFAKNGRAASEBOLNATOT
NIATLGGNAONTNGVAMTDIGGCKNNINDAIKASRNEVKQGNMNVVPTTGANQ
TIVVATADKVAFDVKVGGITIDATTNKISGIAKGDISESTDAVNGSQLYEQQKI
AKSGNYNTLNNRINKVDKDLRAGTAGANAAGLPOAYIPGKSMVAVAAGTYKQNAI
ALGMSRIDNGKVIILKTGNTNRSGDFGASIGAGYOW"
1320 a 688 c 934 g 958 t
BASE COUNT
ORIGIN
```





AKSGDNYNLIINRINKVKDLRAGTAGANAAGLPQAVIPGKSWVAAGTYKQGNAL  
ALGMSRISDNGKVIKLTGNTNSRGDFGASGAGYQW"  
gene complement(6493. .7833)  
/ gene="PM1571"  
CDS complement(6493. .7833)  
/ gene="PM1571"  
/ codon\_start=1  
/ transl\_table=11  
/ product="unknown"  
/ protein\_id="AAK03655.1"  
/ db\_xref="GI:12721964"  
/ translation="WTSSELPNIGFISLGCCKNLVDSERILTELSDGYNIIIPSYENAD  
LVIVTQCGIDSAVQBSLEAIGEALEENGKVIIVTCLGAKEDRIREVHPKVLVETGPH  
SYEAVWQGHVYQFPAYPNFYVNLVPKQGVKLTPIHYAYLKTISEGDKHRCFTCIIPSM  
RGDLSDRSITQVLEAKREIYAGVKEIIVVSQDTSAYALDRSKERQNTVSPWNGMPIK  
NNLISCEQILGKLVWRILHYVYPVHDQLIPLMAEGKILPILDIPIQHASPKTLKA  
KRFSPVEGATLTEMPODVEPEVKEARHFEMQLQOEISAARLOQKVGKFTVTLVDEY  
DESGIARTADAPEIDGVVYIDPNRNVAVKAGQFIEVKITRADAYDLYASLIN"  
gene complement(8029. .8787)  
/ gene="udp"  
/ note="PM1572"  
CDS complement(8029. .8787)  
/ gene="udp"  
/ codon\_start=1  
/ transl\_table=11  
/ product="Udp"  
/ protein\_id="AAK03656.1"  
/ db\_xref="GI:12721965"  
/ translation="MSEVFILGUTKAMLEGATLITPGAPERVEKIAKLLDRPKFLAS  
TRETITGVLGDKAVVYCGTGGSVSTAVBEAQLGVRTLEIGTTGTAQIPHINLV  
GDIUVTGAVRGLDGAISHFAPMEYPAVANEFTNALLYKAATELANQKVYVITAASTD  
FYPCQERYDYSGVYRHFQGSILYKQMODLNMVMEFESATLFTMCSALGLRAGMWAGA  
LVNTQOEIPNEAAVKDIEKNAVEIVVYKAAAYLL"  
gene complement(8810. .9262)  
/ gene="asnC"  
/ note="PM1573"  
CDS complement(8810. .9262)  
/ gene="asnC"  
/ codon\_start=1  
/ transl\_table=11  
/ product="AsnC"  
/ protein\_id="AAK03657.1"  
/ db\_xref="GI:12721966"  
/ translation="MHNIDNQQOILRVLTKDARTPYAEAKNFGVSPGTIHRVVKM  
ROAGIETKVRINERIKGLYDVCCFPIIILKSAKDYDKVTKLEGDFEVVEAYTTGN  
YSIFIKVMTHTIAELHSVLASKIQLIDEIQSTETLISMQNPILHDIAP"  
gene 9420. .10412  
/ gene="asnA"  
/ note="PM1574"  
CDS 9420. .10412  
/ gene="asnA"  
/ codon\_start=1  
/ transl\_table=11  
/ product="AsnA"  
/ protein\_id="AAK03658.1"  
/ db\_xref="GI:12721967"  
/ translation="MKKSFILCOOEISPAKNTETELAEHLGIVEVOGPIILSOVNGNIG  
QNLGSAEKAVQNNVKITDARTFETVHVHSLAYKWKRTIARFNPAQEGILFVMTALRDE  
EDSLDQTHSVYDQNDWEKVIISAEQONLAKVRYAIYAAILITEEAVSKKGLADE  
LPKDIQGFVHSELVQFPNMNDKERENAIKCYEAGVFLIGIGKLSGDKGPHDVRAPDA  
DOWTTPSEGEYGLNGDILVWNPILERAPELISMSGIRYDETALRRQLALATPHNEDRLKE  
DWHQDLVNGRLPLSTGGIGTGRSRLVALLQKKHIGEVQSSWPKWVMEQFDNIL"  
BASE COUNT 2995 a 2124 c 2065 g 3349 t  
ORIGIN



us-09-361-619-10.rge

Thu Sep 13 14:17:47 2001

-----

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: September 12, 2001, 13:39:40 ; Search time 1369.14 Seconds  
(without alignments)  
3183.672 Million cell updates/sec

Title: US-09-361-619-10  
Perfect score: 6942  
Sequence: 1 atgaatacatctataagt.....ttgtgcaggtttcacttt 6942

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues  
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES											
Result No.	Score	Query Match	Length	DB ID	Description						
1	6942	100.0	6942	22	AAF59105	M. catarrhalis les					
2	2939.2	42.3	6159	22	AAF59104	M. catarrhalis str					
3	2852.4	41.1	62909	22	AAF28545	Genomic fragment #					
4	2590	37.3	6144	22	AAF59103	M. catarrhalis str					
5	2590	37.3	6972	22	AAF59102	M. catarrhalis str					
6	2580.6	37.2	6971	17	AAF38740	Moraxella outer me					
7	2580.6	37.2	6973	22	AAF59100	M. catarrhalis str					
8	2485.8	35.8	5979	22	AAF59101	M. catarrhalis str					
9	2484.8	35.8	6259	22	AAF59106	M. catarrhalis M56					
10	147.8	2.1	3030	21	AA92496	Haemophilus influe					
11	146.2	2.1	3036	21	AA92493	Haemophilus influe					

12	141.4	2.0	3354	21	AA92495	Haemophilus influe
13	132.6	1.9	3342	21	AA92498	Haemophilus influe
14	116.6	1.7	720	22	AAF59129	M. catarrhalis str
15	110.6	1.6	936	22	AAF58252	Oligonucleotide D1
16	110.6	1.6	936	22	AAF58254	Oligonucleotide D1
17	110.6	1.6	936	22	AAF58257	Oligonucleotide D1
18	110.6	1.6	936	22	AAF58259	Oligonucleotide D1
19	110.6	1.6	936	22	AAF58262	Oligonucleotide D2
20	110.6	1.6	936	22	AAF58255	Oligonucleotide D1
21	110.6	1.6	936	22	AAF58252	Oligonucleotide D1
22	110.2	1.6	936	22	AAF58254	Oligonucleotide D1
23	110.2	1.6	936	22	AAF58257	Oligonucleotide D1
24	110.2	1.6	936	22	AAF58259	Oligonucleotide D2
25	110.2	1.6	936	22	AAF58262	Oligonucleotide D2
26	110.2	1.6	938	22	AAF58255	Oligonucleotide D1
27	83.6	1.2	2037	17	AA741477	Haemophilus adhesi
28	82.8	1.2	3294	17	AA741475	Haemophilus adhesi
29	82	1.2	2079	21	AA92494	Haemophilus influe
30	80.2	1.2	7486	19	AAV22837	Haemophilus paraga
31	78.6	1.1	8930	19	AAV22834	Haemophilus paraga
32	76.6	1.1	1830121	17	AA742063	Haemophilus influe
33	76.4	1.1	3300	21	AA92497	Haemophilus influe
34	76.4	1.1	7253	21	AA92499	Haemophilus influe
35	76.4	1.1	7291	17	AA741476	Haemophilus adhesi
36	66	1.0	1757	21	AA92500	Haemophilus influe
37	59.6	0.9	2700	21	AA770220	Plasmodium faicipa
38	58.4	0.8	60	17	AA738741	Moraxella outer me
39	58.4	0.8	1797	20	AA85794	DNA encoding a sur
40	56.8	0.8	1797	20	AA85790	DNA encoding a sur
41	55.8	0.8	1779	20	AA99125	DNA encoding N. me
42	55.2	0.8	1770	20	AA85797	DNA encoding a sur
43	55.2	0.8	1776	20	AA99124	DNA encoding N. me
44	55.2	0.8	1776	20	AA85798	DNA encoding a sur
45	55.2	0.8	1776	20	AA85793	DNA encoding a sur

ALIGNMENTS

RESULT 1  
AAF59105  
ID AAF59105 standard; DNA; 6942 BP.  
XX AAF59105;  
AC AC  
XX  
XX 24-APR-2001 (first entry)  
XX  
XX M. catarrhalis les1 200kda gene SEQ ID NO:10.  
DE  
XX Moraxella catarrhalis strain O8; major outer membrane protein;  
KW 200kda outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection; ds.  
XX  
XX Moraxella catarrhalis.  
XX  
XX WO200107619-A1.  
XX  
XX 01-FEB-2001.  
XX  
XX 26-JUL-2000; 2000WO-CA00870.  
XX  
XX 27-JUL-1999; 99US-0361619.  
PR (CONN-) CONNAUGHT LAB LTD.  
XX  
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;  
XX WPI; 2001-159722/16.  
XX P-PSDB; AAB69136.  
XX  
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
XX useful in protective vaccines and for diagnosis

Claim 1; Fig 5A-Y; 247pp; English.

PS The present invention describes an isolated and purified nucleic acid (I)  
 XX that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.  
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and  
 CC can be used in vaccines, (II), and its truncated versions, are used as  
 CC immunogenic compositions and vaccines to protect against M. catarrhalis  
 CC infections, particularly otitis media in humans. (II) is also used as  
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to  
 CC generate Ab. (I) are used for recombinant production of (II) and its  
 CC fragments are used as probes for identifying/cloning 200 kDa protein.  
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
 CC (I) makes possible production of large amount of recombinant immunogens.  
 CC Expression of truncated versions of (II) reduces toxicity of the protein  
 CC towards the Escherichia coli host. The present sequence represents the  
 CC M. catarrhalis les1 200kDa gene, which is given in the exemplification  
 CC of the present invention.

XX  
 SQ Sequence 6942 BP; 2286 A; 1621 C; 1564 G; 1471 T; 0 other;

Query Match 100.0%; Score 6942; DB 22; Length 6942;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 6942; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaatcacatctataaagtcattcttaacaaagccacaggccacattatggccgtggca 60  
 DB 1 atgaatcacatctataaagtcattcttaacaaagccacaggccacattatggccgtggca 60  
 QY 61 gaggcgccaaatccacacagcggaggagtagcagtagtaccgcagcagcaggtggcagc 120  
 DB 61 gaggcgccaaatccacacagcggaggagtagcagtagtaccgcagcagcaggtggcagc 120  
 QY 121 tctctgtcatcccgctgactcgtgtgtgcccagcctgcgtatcctatgtagcagtcgacg 180  
 DB 121 tctctgtcatcccgctgactcgtgtgtgcccagcctgcgtatcctatgtagcagtcgacg 180  
 QY 181 ctcaatggcagtgctatgcataaataatagcaagtcgcatcttggtagcagcagcagc 240  
 DB 181 ctcaatggcagtgctatgcataaataatagcaagtcgcatcttggtagcagcagcagc 240  
 QY 241 aatgacaatgcctggctagcaatgaagcaccatcttctgtagtattgtagtattgtaggca 300  
 DB 241 aatgacaatgcctggctagcaatgaagcaccatcttctgtagtattgtagtattgtaggca 300  
 QY 301 catgccaatcaagctattgctatcgttgtagcagcagcagcagcagcagcagcagc 360  
 DB 301 catgccaatcaagctattgctatcgttgtagcagcagcagcagcagcagcagcagc 360  
 QY 361 aatcagaagcaggttcccacgcccacaaaggttaagagtcctcgcacatcgttgtagtga 420  
 DB 361 aatcagaagcaggttcccacgcccacaaaggttaagagtcctcgcacatcgttgtagtga 420  
 QY 421 ctggctgaggggtagtgcctgattgtagtattgtagtattgtagtattgtagtattgtag 480  
 DB 421 ctggctgaggggtagtgcctgattgtagtattgtagtattgtagtattgtagtattgtag 480  
 QY 481 agcactaaacttaataatccaaatggcttctttagcactcttatttcaaaacacacagta 540  
 DB 481 agcactaaacttaataatccaaatggcttctttagcactcttatttcaaaacacacagta 540  
 QY 541 ttacgccaatacagagactcaaatgttctcagaaatagacgacagcagcagcagcagga 600  
 DB 541 ttacgccaatacagagactcaaatgttctcagaaatagacgacagcagcagcagcagga 600  
 QY 601 caccgcaagtactgagtgaggcagcagcagcagcagcagcagcagcagcagcagcagc 660  
 DB 601 caccgcaagtactgagtgaggcagcagcagcagcagcagcagcagcagcagcagcagc 660  
 QY 661 ggtacacgggtcaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 720  
 DB 661 ggtacacgggtcaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 720

QY 721 gaaaaaggatatacaatcgctattgttctaagtcacaaagctatcaattatggagcacta 780  
 DB 721 gaaaaaggatatacaatcgctattgttctaagtcacaaagctatcaattatggagcacta 780  
 QY 781 gcccttggtgcagatactcgcaggttgatttggtgattacaggtattgccctaggttctct 840  
 DB 781 gcccttggtgcagatactcgcaggttgatttggtgattacaggtattgccctaggttctct 840  
 QY 841 cagatccttaataataataataataataataataataataataataataataataataata 900  
 DB 841 cagatccttaataataataataataataataataataataataataataataataataata 900  
 QY 901 tcaaacataaaatcgtctaaagccaccggaacatgggtttatttccatttgtagtagcact 960  
 DB 901 tcaaacataaaatcgtctaaagccaccggaacatgggtttatttccatttgtagtagcact 960  
 QY 961 atcaacgctaaataatcaatcgtcaggttcaggttatgaggtacacgagcaggtcaatgtg 1020  
 DB 961 atcaacgctaaataatcaatcgtcaggttcaggttatgaggtacacgagcaggtcaatgtg 1020  
 QY 1021 gcacagctaaagcgggtgagaaatctggctaaagcgtcgaatcttcttaaggggtgagat 1080  
 DB 1021 gcacagctaaagcgggtgagaaatctggctaaagcgtcgaatcttcttaaggggtgagat 1080  
 QY 1081 aacgggtactggcgttaagaaacacacggtgagacgtttaaaccattaaaggttgtagagacc 1140  
 DB 1081 aacgggtactggcgttaagaaacacacggtgagacgtttaaaccattaaaggttgtagagacc 1140  
 QY 1141 caagcgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1200  
 DB 1141 caagcgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1200  
 QY 1201 ctgaaagcttaaaacttgctaaacacacacacacacacacacacacacacacacacacac 1260  
 DB 1201 ctgaaagcttaaaacttgctaaacacacacacacacacacacacacacacacacacacac 1260  
 QY 1261 accgccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1320  
 DB 1261 accgccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1320  
 QY 1321 ggtttaaacctttacccac 1380  
 DB 1321 ggtttaaacctttacccac 1380  
 QY 1381 gggcttaagtttactgataattcttaatacagcagcagcagcagcagcagcagcagcagc 1440  
 DB 1381 gggcttaagtttactgataattcttaatacagcagcagcagcagcagcagcagcagcagc 1440  
 QY 1441 gataaaattggttttagcaataaagcgtgtacagttgatgaaacacacacacacacacac 1500  
 DB 1441 gataaaattggttttagcaataaagcgtgtacagttgatgaaacacacacacacacacac 1500  
 QY 1501 aaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1560  
 DB 1501 aaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1560  
 QY 1561 accattggcggtagcaataaacaacacacacacacacacacacacacacacacacacacac 1620  
 DB 1561 accattggcggtagcaataaacaacacacacacacacacacacacacacacacacacacac 1620  
 QY 1621 gtagaattgtagtatacgaatccgcaaaatccgcaaaatccgcaaaatccgcaaaatccg 1680  
 DB 1621 gtagaattgtagtatacgaatccgcaaaatccgcaaaatccgcaaaatccgcaaaatccg 1680  
 QY 1681 attggcttggctgtagtctgtagtctgtagtctgtagtctgtagtctgtagtctgtagtct 1740  
 DB 1681 attggcttggctgtagtctgtagtctgtagtctgtagtctgtagtctgtagtctgtagtct 1740  
 QY 1741 caacttcaagtggtggtgttaaaattacaaagacagtggtggttaattgaagtgatcaaa 1800  
 DB 1741 caacttcaagtggtggtgtgttaaaattacaaagacagtggtggttaattgaagtgatcaaa 1800  
 QY 1801 aagatcggttaagttaaaagtcgaacggagcagcagcagcagcagcagcagcagcagcagc 1860

Db 1801 |||||aaagtcagtaagttaaaagtcgaacgagcagataccgagtcagtcacatttaaacagctt 1860  
QY 1861 aaacaagtcacaagaacccgacgctgcccatacaagcttctctattctgatgaaaaa 1920  
Db 1861 aaacagtcacaagaacccgacgctgcccatacaagcttctctattctgatgaaaaa 1920  
QY 1921 ggtcagaattacagattagtaacttgattcttaagttaatacccccataactcttgag 1980  
Db 1921 ggtcagaattacagattagtaacttgattcttaagttaatacccccataactcttgag 1980  
QY 1981 accatcacctttcaggtgaaaacgcatcagtcagtcagcaatgacatagccaaaggtaaa 2040  
Db 1981 accatcacctttcaggtgaaaacgcatcagtcagtcagcaatgacatagccaaaggtaaa 2040  
QY 2041 gtcaaaagtgtgattgacccaaatcaatggctctcacacgcttaagtcagccgtgggtagc 2100  
Db 2041 gtcaaaagtgtgattgacccaaatcaatggctctcacacgcttaagtcagccgtgggtagc 2100  
QY 2101 gataaagatggtaaaaactcaattggttattgagcaagtgctgagcgttaacgacacaaa 2160  
Db 2101 gataaagatggtaaaaactcaattggttattgagcaagtgctgagcgttaacgacacaaa 2160  
QY 2161 acatcatagagattgtcccaaaactgcctagcattaccatgcaggtgctgacgc 2220  
Db 2161 acatcatagagattgtcccaaaactgcctagcattaccatgcaggtgctgacgc 2220  
QY 2221 accacagaacagggcaatacaatcacacgagcaggaagacaaaatccaaagcccgatc 2280  
Db 2221 accacagaacagggcaatacaatcacacgagcaggaagacaaaatccaaagcccgatc 2280  
QY 2281 ggtgatatataacagctttaacctaaaaataataagcaactcgttgctttgtc 2340  
Db 2281 ggtgatatataacagctttaacctaaaaataataagcaactcgttgctttgtc 2340  
QY 2341 tcaactataacactgtgactttatogatggcaatggccaccgccctgaaggttaacttac 2400  
Db 2341 tcaactataacactgtgactttatogatggcaatggccaccgccctgaaggttaacttac 2400  
QY 2401 gatgaaccacaatacaaacccagtaagtaacttatgatctcaatgtgagtgagaaacacatt 2460  
Db 2401 gatgaaccacaatacaaacccagtaagtaacttatgatctcaatgtgagtgagaaacacatt 2460  
QY 2461 gaactcacggcgataatggtcagagacaaaacaaaattggcgtaaaaacccacactgacc 2520  
Db 2461 gaactcacggcgataatggtcagagacaaaacaaaattggcgtaaaaacccacactgacc 2520  
QY 2521 acaacaaatgctaagttaaagcaacaaactttagtaacaccgataacgctgacctgtt 2580  
Db 2521 acaacaaatgctaagttaaagcaacaaactttagtaacaccgataacgctgacctgtt 2580  
QY 2581 aacgccaagaacatcgccgaaaatctaaacacccctagccaaggaatactcacaccacaaa 2640  
Db 2581 aacgccaagaacatcgccgaaaatctaaacacccctagccaaggaatactcacaccacaaa 2640  
QY 2641 ggcacagcagacacccgctctacaacactttaagtcacaaaagacggtgcaactgtac 2700  
Db 2641 ggcacagcagacacccgctctacaacactttaagtcacaaaagacggtgcaactgtac 2700  
QY 2701 gaaacataccgtgggttaaagatggtacacaaaagcggaagaccgctcaacactctaaaa 2760  
Db 2701 gaaacataccgtgggttaaagatggtacacaaaagcggaagaccgctcaacactctaaaa 2760  
QY 2761 ctcaaaagtgaacacggtctaaacggttgctaccaaataaagatggttacgtttacgttggc 2820  
Db 2761 ctcaaaagtgaacacggtctaaacggttgctaccaaataaagatggttacgtttacgttggc 2820  
QY 2821 attaacacccaaagcgtcttaagccggcagacgacaccactctaaacaaagatggcttg 2880  
Db 2821 attaacacccaaagcgtcttaagccggcagacgacaccactctaaacaaagatggcttg 2880  
QY 2881 tctattaaaaaccccgctagttaacgaacaaaatccaaagtcggtgctgagtgaaagttt 2940  
Db |||||

Db 2881 tctattaaaaaccccgctagttaacgaacaaaatccaaagtcggtgctgagtgaaagttt 2940  
QY 2941 gccaaagttgataagggtaattcaagcactggcattgatggcacaagccgtatcaccaaa 3000  
Db 2941 gccaaagttgataagggtaattcaagcactggcattgatggcacaagccgtatcaccaaa 3000  
QY 3001 gatcaaatggctttactggggctaaatggctcacttgataccaccacaaacccacctaacc 3060  
Db 3001 gatcaaatggctttactggggctaaatggctcacttgataccaccacaaacccacctaacc 3060  
QY 3061 aaagacaaagttaaaagtggtggaattgaaattaccaaacactggcattaaacgaagtggt 3120  
Db 3061 aaagacaaagttaaaagtggtggaattgaaattaccaaacactggcattaaacgaagtggt 3120  
QY 3121 aaaaagattaccaacattcaatcaggtgatattaccacaaacagcaatgatgctgtgaca 3180  
Db 3121 aaaaagattaccaacattcaatcaggtgatattaccacaaacagcaatgatgctgtgaca 3180  
QY 3181 ggcggtggtgttatgatttaaaaaacccgaacttgaaagcaaaaatacaacagtcgtctaaa 3240  
Db 3181 ggcggtggtgttatgatttaaaaaacccgaacttgaaagcaaaaatacaacagtcgtctaaa 3240  
QY 3241 acagcacaaaacattacacgaattctcagtagcagatgaacaaagtaatacactttacg 3300  
Db 3241 acagcacaaaacattacacgaattctcagtagcagatgaacaaagtaatacactttacg 3300  
QY 3301 gttagtaaccttactccagttatgacacctcaagacctctgatgcatcacactttgca 3360  
Db 3301 gttagtaaccttactccagttatgacacctcaagacctctgatgcatcacactttgca 3360  
QY 3361 ggtgaaaacggcattaccacaaaggttaataaaagttggtgctggtgggcattgaccaa 3420  
Db 3361 ggtgaaaacggcattaccacaaaggttaataaaagttggtgctggtgggcattgaccaa 3420  
QY 3421 accaaaggttaaccacgactaaagtcagcgtgggttaataataatgacaaaggtatgctc 3480  
Db 3421 accaaaggttaaccacgactaaagtcagcgtgggttaataataatgacaaaggtatgctc 3480  
QY 3481 attgacagtaaaagatggtcaaaaataccatcacaggaactaaagcaacactctagctaaagt 3540  
Db 3481 attgacagtaaaagatggtcaaaaataccatcacaggaactaaagcaacactctagctaaagt 3540  
QY 3541 accaatgattggtcagagacacgactaagccaaggggttgccaatgacacgcgacaaaacc 3600  
Db 3541 accaatgattggtcagagacacgactaagccaaggggttgccaatgacacgcgacaaaacc 3600  
QY 3601 cgtgccgacagcattggtgatgcttaaaacgacagcgtttaaacttgcaagggcaatgggtgaa 3660  
Db 3601 cgtgccgacagcattggtgatgcttaaaacgacagcgtttaaacttgcaagggcaatgggtgaa 3660  
QY 3661 gcggttgacttctcctcacttatgacacttgacactttatcgatggcaatggccacacc 3720  
Db 3661 gcggttgacttctcctcacttatgacacttgacactttatcgatggcaatggccacacc 3720  
QY 3721 gctaagtgacctatgatgacacaaagcaaaacccagtaaaagtggcttatgagtcaatgtg 3780  
Db 3721 gctaagtgacctatgatgacacaaagcaaaacccagtaaaagtggcttatgagtcaatgtg 3780  
QY 3781 gataataaaacattgaagtgaacagtgataaaaaacttggcgtcaaaaacccacacactg 3840  
Db 3781 gataataaaacattgaagtgaacagtgataaaaaacttggcgtcaaaaacccacacactg 3840  
QY 3841 accaaaacaagtgctaaatggttaatgcaaccaaatttagtgcgcgcatggcgatggccctt 3900  
Db 3841 accaaaacaagtgctaaatggttaatgcaaccaaatttagtgcgcgcatggcgatggccctt 3900  
QY 3901 gttaaagccagtgatgacgacacccatctaaataaccttggctggcgacatcccaacgcgc 3960  
Db 3901 gttaaagccagtgatgacgacacccatctaaataaccttggctggcgacatcccaacgcgc 3960  
QY 3961 aaaggggcaagcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4020  
Db aaaggggcaagcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4020

QY 4021 atctatgacagtaccgataaagaagtactatcatcaagtcaatgacaagggtcaagtggacaaa 4080  
|||||  
Db 4021 atctatgacagtaccgataaagaagtactatcatcaagtcaatgacaagggtcaagtggacaaa 4080  
QY 4081 acaaaagaagtggccaaagacaaactgctgcccacgcccacacccacagatggcacattg 4140  
|||||  
Db 4081 acaaaagaagtggccaaagacaaactgctgcccacgcccacacccacagatggcacattg 4140  
QY 4141 gctcaaatgaaatgtaaaatcagctcatcaacaaagagcaagtaaaatgatgccaaataaaaag 4200  
|||||  
Db 4141 gctcaaatgaaatgtaaaatcagctcatcaacaaagagcaagtaaaatgatgccaaataaaaag 4200  
QY 4201 caaggtcatcaatgaagacacacccctttatcaaaagggcttgaaaacgcgcacaaagacac 4260  
|||||  
Db 4201 caaggtcatcaatgaagacacacccctttatcaaaagggcttgaaaacgcgcacaaagacac 4260  
QY 4261 aaaaacaaacgcgcagcagtaactggtgatttaaatgccgttgcccaaacacgcgtg 4320  
|||||  
Db 4261 aaaaacaaacgcgcagcagtaactggtgatttaaatgccgttgcccaaacacgcgtg 4320  
QY 4321 acctttgcaggggatatacaggcacaacgcgctaaaaactggggcgagactttgacctataaa 4380  
|||||  
Db 4321 acctttgcaggggatatacaggcacaacgcgctaaaaactggggcgagactttgacctataaa 4380  
QY 4381 ggtgggcaaacagacacccaataagctaaacccgataataacatcggtggtgtagcaggtact 4440  
|||||  
Db 4381 ggtgggcaaacagacacccaataagctaaacccgataataacatcggtggtgtagcaggtact 4440  
QY 4441 gatggcttcaactgcaaaacttgccaaagacacccaatacttaacagcgttaatgcaggt 4500  
|||||  
Db 4441 gatggcttcaactgcaaaacttgccaaagacacccaatacttaacagcgttaatgcaggt 4500  
QY 4501 ggcacacgaatgatgaaaaagacatctctttgttagacgcaaacggtcaagccaaaagca 4560  
|||||  
Db 4501 ggcacacgaatgatgaaaaagacatctctttgttagacgcaaacggtcaagccaaaagca 4560  
QY 4561 aacacccctgcttaagtccaatgggctggacctgggtggcaaacgcgcacgttaacatc 4620  
|||||  
Db 4561 aacacccctgcttaagtccaatgggctggacctgggtggcaaacgcgcacgttaacatc 4620  
QY 4621 ggtgcagctgtgatgaacgatgcggtgaactttaagcagtttaagtggccaaa 4680  
|||||  
Db 4621 ggtgcagctgtgatgaacgatgcggtgaactttaagcagtttaagtggccaaa 4680  
QY 4681 acggtcaaaccttaaacacacaaagtaactcaggtgcgtcattacccctttggttaacc 4740  
|||||  
Db 4681 acggtcaaaccttaaacacacaaagtaactcaggtgcgtcattacccctttggttaacc 4740  
QY 4741 gatgccaatggcaagcccaataaaatggcacccgatggcaagcccaaaaagccatcaagggc 4800  
|||||  
Db 4741 gatgccaatggcaagcccaataaaatggcacccgatggcaagcccaaaaagccatcaagggc 4800  
QY 4801 gccgatgtaataactatcacgccaacgcgaacgcggtacacttggaacaaagatggcaag 4860  
|||||  
Db 4801 gccgatgtaataactatcacgccaacgcgaacgcggtacacttggaacaaagatggcaag 4860  
QY 4861 cccatcacccgatgggacaaacttgccaaatctggcagctcatggcacaaccccttgatga 4920  
|||||  
Db 4861 cccatcacccgatgggacaaacttgccaaatctggcagctcatggcacaaccccttgatga 4920  
QY 4921 ggtcatcaagtgggtggcaagcctaggcggaactcagatgcacccaataacacatc 4980  
|||||  
Db 4921 ggtcatcaagtgggtggcaagcctaggcggaactcagatgcacccaataacacatc 4980  
QY 4981 aagtcacactttgccaacaaatggacacacacacacacaggttaatgccaatgagggcaagcc 5040  
|||||  
Db 4981 aagtcacactttgccaacaaatggacacacacacacacaggttaatgccaatgagggcaagcc 5040  
QY 5041 caaagtgtcccagcctcatcagcagcagcagcaaaagtaatgctgcagtgtaaaagatgtg 5100  
|||||  
Db 5041 caaagtgtcccagcctcatcagcagcagcagcaaaagtaatgctgcagtgtaaaagatgtg 5100

QY 5101 ctaaaatgtaggcttttaacttgcagacccaatacaaatggaacttctgttcaaaagcctat 5160  
|||||  
Db 5101 ctaaaatgtaggcttttaacttgcagacccaatacaaatggaacttctgttcaaaagcctat 5160  
QY 5161 gatccgctcaactttgtcaatggttacaggtgcgacatcacaaagcgtgcgtagtgctgat 5220  
|||||  
Db 5161 gatccgctcaactttgtcaatggttacaggtgcgacatcacaaagcgtgcgtagtgctgat 5220  
QY 5221 ggcacgatgagtaacatcacctcaacacgcgcttagcagcgaccgatgatggcaat 5280  
|||||  
Db 5221 ggcacgatgagtaacatcacctcaacacgcgcttagcagcgaccgatgatggcaat 5280  
QY 5281 gttcttatcaaaagccaaagatggttaagttctcaaaagcagacacatcgcacaaagcgc 5340  
|||||  
Db 5281 gttcttatcaaaagccaaagatggttaagttctcaaaagcagacacatcgcacaaagcgc 5340  
QY 5341 tcactaaaacagcgaacaaatcagccagtgatgccaaaactccaaactggtctlaagccttgtt 5400  
|||||  
Db 5341 tcactaaaacagcgaacaaatcagccagtgatgccaaaactccaaactggtctlaagccttgtt 5400  
QY 5401 aaccccaatgctggttaaaagcagtcacaggcgatgagctggctcttaataacttatcaaaa 5460  
|||||  
Db 5401 aaccccaatgctggttaaaagcagtcacaggcgatgagctggctcttaataacttatcaaaa 5460  
QY 5461 gcggtatttaataccaaagatggtacaaactactaccacagtaagctctgctggtcagtcagt 5520  
|||||  
Db 5461 gcggtatttaataccaaagatggtacaaactactaccacagtaagctctgctggtcagtcagt 5520  
QY 5521 atccaaagcgaagataacagcagcatcaccttaagcgaagatggctgtaagtgagggcgt 5580  
|||||  
Db 5521 atccaaagcgaagataacagcagcatcaccttaagcgaagatggctgtaagtgagggcgt 5580  
QY 5581 aaggtcatcaacaaatgtgggttaaaagcacaacaaagacacacacgcgtgcgaatgtacaacag 5640  
|||||  
Db 5581 aaggtcatcaacaaatgtgggttaaaagcacaacaaagacacacacgcgtgcgaatgtacaacag 5640  
QY 5641 ttaaacgaagtacgcaacttgttggtctgtggttaagtgcgttaataatgataacgctgacggc 5700  
|||||  
Db 5641 ttaaacgaagtacgcaacttgttggtctgtggttaagtgcgttaataatgataacgctgacggc 5700  
QY 5701 aatcaggtaaacatgtgcgacatcaaaaaagacccaaatcaggttcatctatcaaccgc 5760  
|||||  
Db 5701 aatcaggtaaacatgtgcgacatcaaaaaagacccaaatcaggttcatctatcaaccgc 5760  
QY 5761 actgtcatcaaaagcagcgacggtacttggcggtaaagttaataacgataccgcaaaaactt 5820  
|||||  
Db 5761 actgtcatcaaaagcagcgacggtacttggcggtaaagttaataacgataccgcaaaaactt 5820  
QY 5821 gccactggtggtgtgtcaagtggcggtgataaaagcgcgaacgctaaagcggttaagc 5880  
|||||  
Db 5821 gccactggtggtgtgtcaagtggcggtgataaaagcgcgaacgctaaagcggttaagc 5880  
QY 5881 aatggttgggtcaaaacccaaagatggcagcaaaaagccctgctgcacacttataac 5940  
|||||  
Db 5881 aatggttgggtcaaaacccaaagatggcagcaaaaagccctgctgcacacttataac 5940  
QY 5941 gccgaggttcagacaaactatttgacaaacaccccgcaagaagccattgacagataaat 6000  
|||||  
Db 5941 gccgaggttcagacaaactatttgacaaacaccccgcaagaagccattgacagataaat 6000  
QY 6001 gaacaaaggtatccgctcttctccatgtccaacgatggcaatcgaagacccctgtggtacaaaggg 6060  
|||||  
Db 6001 gaacaaaggtatccgctcttctccatgtccaacgatggcaatcgaagacccctgtggtacaaaggg 6060  
QY 6061 cgtaacgcatgtgactcaagtgcctcaggcaagcactcagtgcgatagagtttccagggcc 6120  
|||||  
Db 6061 cgtaacgcatgtgactcaagtgcctcaggcaagcactcagtgcgatagagtttccagggcc 6120  
QY 6121 aagcgagatggtgaagccgcgttgccatagcgagacaaaccccaacagcaggaaccaatcc 6180  
|||||  
Db 6121 aagcgagatggtgaagccgcgttgccatagcgagacaaaccccaacagcaggaaccaatcc 6180  
QY 6181 atcgccatcggtgataacgcacaaagccacggggtgcataatcccatcgccatcggtacagggc 6240



```

Db 6181 atcgccatcggtgatacgcacacgacgagcgagcgatcaatcccatcgccatcggtacagcg 6240
Qy 6241 aatggtgtaacagtaagcactctggcgccatcgcgacccaagcactgttaagcgctgat 6300
Db 6241 aatggtgtaacagtaagcactctggcgccatcgcgacccaagcactgttaagcgctgat 6300
Qy 6301 aacagttacagtggtggtgtaataacacacagctttatgatcgccactcagacgatgtcttt 6360
Db 6301 aacagttacagtggtggtgtaataacacacagctttatgatcgccactcagacgatgtcttt 6360
Qy 6361 ggtggtggcaataacatcacgcgtgacggaagtaactcgttgcttaggttcaaaactct 6420
Db 6361 ggtggtggcaataacatcacgcgtgacggaagtaactcgttgcttaggttcaaaactct 6420
Qy 6421 gccatcagtgagggacacacgacgagcgacacagcgcaaaaaatctgacggcgacagcggt 6480
Db 6421 gccatcagtgagggacacacgacgagcgacacagcgcaaaaaatctgacggcgacagcggt 6480
Qy 6481 acaaccacacagcagtgcaacaggttacggtttaaaggctttgctgacaaaacggcggtt 6540
Db 6481 acaaccacacagcagtgcaacaggttacggtttaaaggctttgctgacaaaacggcggtt 6540
Qy 6541 ggtgcggtctccgtgggtgctcaggtgctgaacgcggtatccaaaatgtggcagcaggt 6600
Db 6541 ggtgcggtctccgtgggtgctcaggtgctgaacgcggtatccaaaatgtggcagcaggt 6600
Qy 6601 gaggtcagtgccacacagcagcgatcggttcaatggttagcaggtgtgtaaaagccacccaa 6660
Db 6601 gaggtcagtgccacacagcagcgatcggttcaatggttagcaggtgtgtaaaagccacccaa 6660
Qy 6661 ggcattgccaaacgcaacaaatgagcttgacatcgatccacaaaacgaaaataaagcc 6720
Db 6661 ggcattgccaaacgcaacaaatgagcttgacatcgatccacaaaacgaaaataaagcc 6720
Qy 6721 aatgcagggatttcacagcagtgccgagtgccatgcacaaagcctacattcctggc 6780
Db 6721 aatgcagggatttcacagcagtgccgagtgccatgcacaaagcctacattcctggc 6780
Qy 6781 agatccatgattaccgggggtattgccacccacaaacgggtcaaggctgcggtggcagtgga 6840
Db 6781 agatccatgattaccgggggtattgccacccacaaacgggtcaaggctgcggtggcagtgga 6840
Qy 6841 ctgtcgaagctgctcgataatggtcaatggtggtatttaaaatcaatggttcagcgcgatacc 6900
Db 6841 ctgtcgaagctgctcgataatggtcaatggtggtatttaaaatcaatggttcagcgcgatacc 6900
Qy 6901 caaggccatgtagggcgcgagtggtgagcaggttttcaacttt 6942
Db 6901 caaggccatgtagggcgcgagtggtgagcaggttttcaacttt 6942

```

## RESULT 2

AAF59104

ID AAF59104 standard; DNA; 6159 BP.

XX

AC AAF59104;

XX

XX 24-APR-2001 (first entry)

XX

XX M. catarrhalis strain Q8 200kDa gene SEQ ID NO:8.

XX

XX Moraxella catarrhalis strain Q8; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KW otitis media; detection; ds.

XX

OS Moraxella catarrhalis.

XX

XX WO200107619-A1.

XX

XX 01-FEB-2001..

XX

XX 26-JUL-2000; 2000WO-CA00870.

XX

```

XX 27-JUL-1999; 99US-0361619.
XX (CONN-) CONNAUGHT LAB LTD.
XX PA
XX PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX DR WPI; 2001-159722/16.
XX P-PSDB; AAB69135.
XX PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX DR useful in protective vaccines and for diagnosis
XX Claim 1; Fig 4A-V; 247pp; English.
XX
XX The present invention describes an isolated and purified nucleic acid (I)
XX CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
XX CC The 200 kDa outer membrane protein (II) has antibacterial activity and
XX CC can be used in vaccines. (II), and its truncated versions, are used as
XX CC immunogenic compositions and vaccines to protect against M. catarrhalis
XX CC infections, particularly otitis media in humans. (II) is also used as
XX CC antigen in immunoassays for detecting specific antibodies (Ab), and to
XX CC generate Ab. (I) are used for recombinant production of (II) and its
XX CC fragments are used as probes for identifying/cloning 200 kDa protein
XX CC genes from other strains, and for diagnostic detection of M. catarrhalis.
XX CC (I) makes possible production of large amount of recombinant immunogens.
XX CC Expression of truncated versions of (II) reduces toxicity of the protein
XX CC towards the Escherichia coli host. The present sequence represents the
XX CC M. catarrhalis strain Q8 200kDa gene, which is given in the
XX CC exemplification of the present invention.
XX
XX Sequence 6159 BP; 2035 A; 1386 C; 1385 G; 1353 T; 0 other;

```

Query Match 42.3%; Score 2939, 2; DB 22; Length 6159;

Best Local Similarity 70.2%; Pred. No. 0;

Matches 5085; Conservative 0; Mismatches 768; Indels 1395; Gaps 29;

Qy 1 atgaatcacatctataaagtcacatctttaacaaagccacagggcacatttatggcgtggca 60

Db 1 atgaatcacatctataaagtcacatctttaacaaagccacagggcacatttatggcgtggcg 60

Qy 61 gagtgcgcaaaatccacacagcgaggagtagcagtagtaccgcagcagcagcagcagc 120

Db 61 gaatatgccaaatccacacagctacgggggggtagctgtctacagggcgaagtggcagt 120

Qy 121 tctctgtcatccgcctgacacgctgttgccacgctcctatcctagtgatcggtgcgacg 180

Db 121 gtacgcaacttaagcttgcccgctattgcgcgctcctcctcgtcgtcgtgcgacg 180

Qy 181 ctcaatggcagtgcttatgct---caaaataatagcaagatcgcatattggtaccacagcg 237

Db 181 ctcaatggcagtgcttatgctcaacaaattactaccaagatcgaaattggtcaacaacac 240

Qy 238 aa-----caatgacaatgctcgtcgctcgtcgtcgtcgtcgtcgtcgtcgtcgtt 282

Db 241 aagataaacacacgcgtgaaagcgatgcccctagcagcagcagcagcagcagcagcagc 300

Qy 283 ggtagtcttgcgaagcagacatgccaatcaagctattgctatcgtggtagcaaacacagat 342

Db 301 ggtagtcttgcgaagcagacagcgtctcaagctattgctatcgtcgtcgtcgtcgtcgt 360

Qy 343 cctcgtcaatcaacgcgttaatacagaagcgaggttccacacgcaaaaggtaagagtcctac 402

Db 361 cc---taataatggtagtaatggttaagtgaggttccacacgcaaaaggtaagagtcctac 417

Qy 403 gccatcggtggtgtagtactggtcgtgaggtgtagcctcgtcgtcgtcgtcgtcgtcgtcgt 462

Db 418 gccatcggtggtgtagtactggtcgtgaggtgtagcctcgtcgtcgtcgtcgtcgtcgtcgt 477

Qy 463 ttatatttgataggaatagcactaactctaaatataatcgaatggtctctttagcactctt 522

Db 478 ttatatttgataggaatagcactaactctaaatataatcgaatggtctctttagcactctt 525





QY 4615 aacatcgggtgacgtgttgatgataacgatcggtgaaactttaagcagtttaataagatt 4674  
 Db 4807 ----- 4806  
 QY 4675 gccaaaacggtcaacaaactaaacaaacaaagtaactcaggtgcgtcattaccctttgtg 4734  
 Db 4807 ----- 4806  
 QY 4735 gtaaccgatgccatggcaagcccatcaatggcaccgatggcaagcccaaaaagccatc 4794  
 Db 4807 ----- 4806  
 QY 4795 aagggcgccgatggtaataactatcacgccaaacgccaacggtacgtgtggacaagaat 4854  
 Db 4807 ----- 4806  
 QY 4855 ggcaagcccatcaccgatgcggaacaaacttgccaatctggcagctcatggcaaacccctt 4914  
 Db 4807 ----- 4806  
 QY 4915 gatgcaggtcatcaagtgggcaagccttagcgcgcaactcagatgccatcacccctaacc 4974  
 Db 4807 ----- 4806  
 QY 4975 aacatcaagtccaacttggccacaaattgacacacacacacaggttaatgccaatgcaggg 5034  
 Db 4807 ----- 4806  
 QY 5035 caagcccaaaagtctgccagccttatcagcagcacagcaagcaaaagtaatgctgcagttgtcaaa 5094  
 Db 4807 ----- 4806  
 QY 5095 gatgtcctaagttaggctttaacttgcagaccacaaatcaaatcaagtggacttggtaaaa 5154  
 Db 4807 ----- 4806  
 QY 5155 gcctatgataccgttcaactttgtcaattggtacaggtgccgacatcacaaagcttgogtagt 5214  
 Db 4807 ----- 4806  
 QY 5215 gctgatggcacgatgagtaacatacacogtcaacacgccttagcagcagcagatgatgat 5274  
 Db 4807 ----- 4806  
 QY 5275 ggcaatgtgottatcaaaagccaaagatggtaagttctacaaagcagacagacacotcatgcca 5334  
 Db 4807 ----- 4806  
 QY 5335 aacggtctactaaagcagcgcaaatcagccagtgatgccaaaaactccaactggtcttaagc 5394  
 Db 4807 ----- 4806  
 QY 5395 ctgttaaccccaatgctggtaagggcagttacagggcagtgatgcagtggtgctcttaataacta 5454  
 Db 4807 ----- 4806  
 QY 5455 tcaaaaagcggattttaaatccaaagatggtaacactactaccacagtaagctctgtatggc 5514  
 Db 4807 ----- 4806  
 QY 5515 atcagatccaaagcaaaagataacagcagcatcacccctaaagcaaaagatgggctgaatgta 5574  
 Db 4807 ----- 4806  
 QY 5575 ggcgtaaggttcacagcaaatgggttaagggcaaaaagacacccgacgtgcacaaatgta 5634  
 Db 4807 ----- 4806  
 QY 5635 caacagttaaacgaagtacgaacattgtgggtcttggtaatgctggtaataatgataacgct 5694  
 Db 4861 caacagttaaacgaagtaacgaactgttgggtct-----tggcaatgataacgct 4911

QY 5695 gacggcaatcaggttaaacattgccgacatacaaaaaagaccccaattcaggttcatcatct 5754  
 Db 4912 gacggcaatcaggttaaacattgccgacatacaaaaaagaccccaattcaggttcatcatct 4971  
 QY 5755 aaccgcactgtcatcaaaagcagcacggttacttggcggtaaaaggttaataacgataccgaa 5814  
 Db 4972 aaccgcactgtcatcaaaagcagcacggttacttggcggtaaaaggttaataacgataccgaa 5031  
 QY 5815 aaacttgcacactgttgggtgataaagtggcggtgataaaagacgggcaacgtacacgcgcat 5874  
 Db 5032 aaacttgcacactgttgggtgataaagtggcggtgataaaagacgggcaacgtacacgcgcat 5091  
 QY 5875 ttaagcaatgtttgggtgataaaaccccaaaagataggcagcaaaaaaacccctgctcgcaact 5934  
 Db 5092 ttaagcaatgtttgggtgataaaaccccaaaagataggcagcaaaaaaacccctgctcgcaact 5151  
 QY 5935 tataacgcccaggtcagacccaactatttgacaaacaccccccagaaagccattgacaga 5994  
 Db 5152 tataacgcccaggtcagacccaactatttgacaaacaccccccagaaagccattgacaga 5211  
 QY 5995 ataaatgaacaaaggtatccgcttcttccatgttcaacagatggcgaatcaaaagcctgtggta 6054  
 Db 5212 ataaatgaacaaaggtatccgcttcttccatgttcaacagatggcgaatcaaaagcctgtggta 5271  
 QY 6055 caagggcgtaacggcattgactcaagtgcctcaggaagcactcagtgggcgataggtttc 6114  
 Db 5272 caagggcgtaacggcattgactcaagtgcctcaggaagcactcagtgggcgataggtttc 5331  
 QY 6115 caggccaaagcagatggtgaaagccgcctgtccatagggcagacaaaccccaagcaggaac 6174  
 Db 5332 caggccaaagcagatggtgaaagccgcctgtccatagggcagacaaaccccaagcaggaac 5391  
 QY 6175 caatccatcgccatcggtgataacgacacagccacgggacgaatcaatccatcgccatcggt 6234  
 Db 5392 caatccatcgccatcggtgataacgacacagccacgggacgaatcaatccatcgccatcggt 5451  
 QY 6235 acaggcaatgtggttaacaggttaagcaactgtgctccatcgccgaccccaagcactgttaag 6294  
 Db 5452 acaggcaatgtggttaacaggttaagcaactgtgctccatcgccgaccccaagcactgttaag 5511  
 QY 6295 gctgataacagttacagtggtgggttaataacacacagtttatcgatgccactcagacccgat 6354  
 Db 5512 gctgataacagttacagtggtgggttaataacacacagtttatcgatgccactcagacccgat 5571  
 QY 6355 gctcttgggtgggcaataacatcacctgacccgaaagttaactcgggttgccttaggttca 6414  
 Db 5572 gctcttgggtgggcaataacatcacctgacccgaaagttaactcgggttgccttaggttca 5631  
 QY 6415 aactctgcaatcagtgccagggcacacacagcagccacacagcagccacacacacacacacac 6474  
 Db 5632 aactctgcaatcagtgccagggcacacacagcagccacacacagcagccacacacacacacac 5691  
 QY 6475 gcaggtacaaacacacacacacagcaggtgcaacaggtacgggttaaaagccttgcctggacaaacg 6534  
 Db 5692 gcaggtacaaacacacacacacagcaggtgcaacaggtacgggttaaaagccttgcctggacaaacg 5751  
 QY 6535 gcgggttgggtgcctcctgggtgcctcaggtgcctcaggtgcctcaggtgcctcaggtgcctcaggt 6594  
 Db 5752 gcgggttgggtgcctcctgggtgcctcaggtgcctcaggtgcctcaggtgcctcaggtgcctcaggt 5811  
 QY 6595 gcaggtgcaggtgcagtgccac 6654  
 Db 5812 gcaggtgcaggtgcagtgccac 5871  
 QY 6655 acccaaggtcattgccaacgcac 6714  
 Db 5872 acccaaggtcattgccaacgcac 5931  
 QY 6715 aaagcgaatgcaggttgcagcgaatgcagcgaatgcagcgaatgcagcgaatgcagcgaatgcagcga 6774  
 Db 5932 aaagcgaatgcaggttgcagcgaatgcagcgaatgcagcgaatgcagcgaatgcagcgaatgcagcga 5991  
 QY 6775 cctggcagatccatggttacccgggggtatttgccaccccaacacacacacacacacacacacacac 6834

```

Db 5992 ccgggcagatccatggttacccgggggtattgcccaccacacaggtcaagggtcggtggca 6051
Qy 6835 gtgggactgtcgaagctgtcgataaaggtaattgggtatttataaaatcaatgggtcgaagc 6894
Db 6052 gtgggactgtcgaagctgtcgataatggtaattgggtatttataaaatcaatgggtcgaagc 6111
Qy 6895 gatacccaaggccatgtagggggcggcagttggtgtaggggttttcaacttt 6942
Db 6112 gatacccaaggccatgtagggggcggcagttggtgtaggggttttcaacttt 6159

RESULT 3
AAF28545
ID AAF28545 standard; DNA; 62909 BP.
XX
AC AAF28545;
XX
XX 04-APR-2001 (first entry)
DE Genomic fragment #32.
XX
XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.
XX
XX Moraxella catarrhalis.
OS
XX WO200078968-A2.
XX
XX 28-DEC-2000.
XX
XX 16-JUN-2000; 2000WO-US16649.
XX
XX 18-JUN-1999; 99US-0140121.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Lagace RE, Patterson C, Berg KL;
PI WPI; 2001-041427/05.
XX
XX Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids -
XX
XX Claim 1; Page 293-308; 545pp; English.
XX
XX The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
XX Sequence 62909 BP; 18449 A; 13798 C; 12744 G; 17918 T; 0 other;
XX

Query Match 41.1%; Score 2852.4; DB 22; Length 62909;
Best Local Similarity 69.3%; Pred. No. 0;
Matches 4853; Conservative 0; Mismatches 991; Indels 1156; Gaps 25;

Qy 1 atgaatcacatctataaagtcatttttaacaaagccacaggcacatttatggcgtggca 60
Db 42630 atgaatcacatctataaagtcatttttaacaaagccacaggcacatttatggcgtggc 42689
Qy 61 gaatggccaaatcccaagc-gagaggagtagcagtagtacccagcagcagtagggcag 119
Db 42690 gaatggccaaatcccaagcagcagcgggggggtagctgtgctacagggcaagtggcag 42749

```

```

Qy 120 ctctctgtctacgcgcctgactcgtgttggcacgctcgctctatcctagtatcgtgtagc 179
Db 42750 tgatgcactcgaagcttgcgcgtattgcgcgctcgtcgtcgtcgtcgtcgtcgtcgtc 42809
Qy 180 gctcaatggcagtgcttctatctcaaaataatagcaaatgcgcatcttggtaccacagcaa 239
Db 42810 gctcagtggcagtgcttctatctcaagta---tcgcaatggccaacaaacaaacac 42866
Qy 240 caatgacaatgcctcgcgttagcaatgaagcatccatccttgctattggtagtcttgccta 299
Db 42867 aaaaggcagtgccgagggcgaggaagtaatcaatcogtctgctcttggtagaaagctgtgc 42926
Qy 300 acatgccaatcaagctattgctatcggtggtagcaaacccagatcctcgtgtaatacagggc 359
Db 42927 aagtggcagtcgaagccatcgccatcggtggtagccaaacaaatcaaaataa----- 42977
Qy 360 taatcagaaggcaggtgtccacgccaaaaggtaaaagatccatcgccatcggtggtgagt 419
Db 42978 tagtaaatgtaggtgccaacgcaaaaggtaagagtagtccatcgccatcggtggtgagt 43037
Qy 420 actggctgagggtagtcctcgattgccattggttagtgtagtgaacttatattgatatagaa 479
Db 43038 attggctcaggggtcctcgctcgattgccatcggttagtcgtagtgaacttatattgtagaa 43097
Qy 480 tagcactaaacttaaatatccaaatggtctcttcttagcactcttattcacaacccatacagt 539
Db 43098 agg-----ctgtatatacactcccaagagattaaaaaacttattgaaaaacattcaga 43151
Qy 540 attacgcaaaatacga-----gactcaaatggtcttcagaaaaatatagacgacagcagc 593
Db 43152 attaaacagatacgaaaactggatcccaatagtaatacaaaaattattacacaaaagc 43211
Qy 594 agaaggacgcccagtagctcagtgggagccatggcagctatgcaaaagggtcattttgccaa 653
Db 43212 acaaggacgcccagtagtattgcagtgggaaacatggcagcagaggggtcattttgccaa 43271
Qy 654 cgcttggtagcaggttcaaacagctgaaggcaactattccttggcagtaggtcttaccgc 713
Db 43272 cgcttggtagcaggttcaaacagctgaaggcaactattccttggcagtggttaccgc 43331
Qy 714 caaagcgaagaaaggatatacaatcgtattggttcttaatgacaaagctatcaattatgg 773
Db 43332 caaagcgaagaaaggatatacaatcgtattggttcttaatgacaaagctatcaattatgg 43391
Qy 774 agcactagcccttggtgcagatactcagtgattgattgattgattgattgattgattgatt 833
Db 43392 agcagtagccattggtggaaataactaaatctgagttgggtcgaggtgttgccttagtta 43451
Qy 834 tgggttcagatccttaataataataataataataataataataataataataataataataag 893
Db 43452 taattctcagatc---actgctaataataataataataataataataataataataataataag 43508
Qy 894 taatgggtcaaacataaaaactgctcaaaagccacggcaat-----gg 935
Db 43509 tgaacattacacagagcatgaagccacaaaacggcagcgaagaataacagtgga 43568
Qy 936 ttatttccattggttagtagcactatcaagcgtaaaatcatcaatgctcgtgtagtta 995
Db 43569 tataatttccattggttaataatgtagtctcagacgttagaataatccatcctcgtgtaggt 43628
Qy 996 tgaaggataccgtagcgtcgaatgtggcacagctaaaagcgtgagagatctgggctaagcg 1055
Db 43629 tgaatgccacgtagcgttcaatgtggcacagtagaagcgtggtgtaactgggctaagaa 43688
Qy 1056 tcaaatctactttaaagggtgatataacggtactggttaagaaaaaaactgggagagac 1115
Db 43689 gcgtaaaattacttttaagggtgatgacagtactggtcgtcaaaaaggttttcgacgagga 43748
Qy 1116 tttaaccattaaaggtgtagaccacgagcgaagcaagcaagcaagcaagcaagcaagcaagca 1175
Db 43749 gtttaactattaaaggcggcgagcaaatgcgggtgaattacacgacataaataacatcggt 43808

```







QY 5523 ccaaggcaagataaacagcagcatcaccttaagcaaaagatgggctgaatgtaggcggttaa 5582  
DB 47170 -----aa 47171  
QY 5583 ggtcatcagcaatgtgggttaaaggcacaaagacaccgacgctgccaatgtacaacagtt 5642  
DB 47172 ggtcatcagcaatgtgggttaaaggcacaaagacaccgacgctgccaatgtacaacagtt 47231  
QY 5643 aaacgaagtacgcaactgtgtgggtcttggttaattgtctgtaataatgataacgctgacggcaa 5702  
DB 47232 aaacgaagtacgcaactgtgtgggtcttggttaattgtctgtaataatgataacgctgacggcaa 47291  
QY 5703 tcaggttaacattgcgcacatcaaaaagacccaaatcaggttcatcatctaaacgcac 5762  
DB 47292 tcaggttaacattgcgcacatcaaaaagacccaaatcaggttcatcatctaaacgcac 47351  
QY 5763 tgtcatcaaaagcagcagctacttggtggttaaaaggttaataacgataacgaaaaacttgc 5822  
DB 47352 tgtcatcaaaagcagcagctacttggtggttaaaaggttaataacgataacgaaaaacttgc 47411  
QY 5823 cactggtgtgtacaagtggcggtgataaaagacgcaacgctaacgctgaatttaagcaa 5882  
DB 47412 cactggtgtgtacaagtggcggtgataaaagacgcaacgctaacgctgaatttaagcaa 47471  
QY 5883 tgttgggtcacaaccccaaaagatggcgcaaaaagccctgctgcacattataacgc 5942  
DB 47472 tgttgggtcacaaccccaaaagatggcgcaaaaagccctgctgcacattataacgc 47531  
QY 5943 cgcagtcagacacaactatttgacacaacaccccgacagcattgacagataataatga 6002  
DB 47532 cgcagtcagacacaactatttgacacaacaccccgacagcattgacagataataatga 47591  
QY 6003 acaagttaccgttcttcattgcaacgatggcaatcaagacctgtgtgtacaaggcg 6062  
DB 47592 acaagttaccgttcttcattgcaacgatggcaatcaagacctgtgtgtacaaggcg 47651  
QY 6063 taacgcattgactcaagtgcctcaggcaagcactcagtgggcgataggtttccaggccaa 6122  
DB 47652 taacgcattgactcaagtgcctcaggcaagcactcagtgggcgataggtttccaggccaa 47711  
QY 6123 ggcagatgtggaagccgctgtccatagcagacaaacccagcaggcaacacacccat 6182  
DB 47712 ggcagatgtggaagccgctgtccatagcagacaaacccagcaggcaacacacccat 47771  
QY 6183 cgcacatggtgataacgcaaaagccagcggtgatcaatccatccatcgccatcggtacaggcaa 6242  
DB 47772 cgcacatggtgataacgcaaaagccagcggtgatcaatccatccatcgccatcggtacaggcaa 47831  
QY 6243 tgtgttaacaggttaagcactctgtgccatcgccagcccaagcactgttlaaggctgataa 6302  
DB 47832 tgtgtgtagcaggttaagcactctgtgccatcgccagcccaagcactgttlaaggctgataa 47891  
QY 6303 cagttacagttgggttaataaacacagtttatcgatgcacactcagacagcagtgctttgg 6362  
DB 47892 cagttacagttgggttaataaacacagtttatcgatgcacactcagacagcagtgctttgg 47951  
QY 6363 tgtgtggaataaacatccctgacccgaaagttaactcgtgttccttaggttcaaacctcgc 6422  
DB 47952 tgtgtggaataaacatccctgacccgaaagttaactcgtgttccttaggttcaaacctcgc 48011  
QY 6423 catcagtcagggcacacacgagcacacaaagcaaaaatctgacggcacagcaggtac 6482  
DB 48012 catcagtcagggcacacacgagcacacaaagcaaaaatctgacggcacagcaggtac 48071  
QY 6483 aaccacacagcaggtgcaacaggtacggttaaaaggcttgcctggacaaacgctgttg 6542  
DB 48072 aaccacacagcaggtgcaacaggtacggttaaaaggcttgcctggacaaacgctgttg 48131  
QY 6543 tgcggttcagtggttcagctcaggtgctgaacgccttatccaaaatgtgcagcaggtga 6602  
DB 48132 tgcggttcagtggttcagctcaggtgctgaacgccttatccaaaatgtgcagcaggtga 48191  
QY 6603 ggtcagtgccaccagcagcagctgctggtggttaattggtagccagttgtacaagaaccccaagg 6662

DB 48192 ggtcagtcgcccagcagcagctggtcgaatggttagcagttgtacaagcccccag 48251  
QY 6663 cattgccaacgcaacaaatgagcttgaccatcgatccaccacaaacgaaaaataaagccaa 6722  
DB 48252 cattgccaacgcaacaaatgagcttgaccatcgatccaccacaaacgaaaaataaagccaa 48311  
QY 6723 tgcagggatttcacatcagcagctggtggttccatgcccacaaagcctacattcctggcag 6782  
DB 48312 tgcagggatttcacatcagcagctggtggttccatgcccacaaagcctacattcctggcag 48371  
QY 6783 atccatggttacgggggttattgccaccacaaacgctcaaggtgcagtgcaagtggagct 6842  
DB 48372 atccatggttacgggggttattgccaccacaaacgctcaaggtgcagtgcaagtggagct 48431  
QY 6843 gtcgaagctgcgataatggtcaatgggttatataaatacaatggttcagcccgatataccca 6902  
DB 48432 gtcgaagctgcgataatggtcaatgggttatataaatacaatggttcagcccgatataccca 48491  
QY 6903 aggcattgtagggggcgagctggtgcaaggttttcacttt 6942  
DB 48492 aggcattgtagggggcgagctggtgcaaggttttcacttt 48531  
RESULT 4  
AAF59103  
ID AAF59103 standard; DNA; 6144 BP.  
XX AAF59103;  
XX 24-APR-2001 (first entry)  
DE M. catarrhalis strain 4223 genomic 200kDa coding sequence SEQ ID NO:6.  
KW Moraxella catarrhalis strain 4223; major outer membrane protein;  
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection; ds.  
XX Moraxella catarrhalis.  
XX WO200107619-A1.  
XX 01-FEB-2001.  
XX 26-JUL-2000; 2000WO-CA00870.  
XX 27-JUL-1999; 99US-0361619.  
XX (CONN-) CONNAUGHT LAB LTD.  
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;  
XX WPI; 2001-159722/16.  
XX P-PSDB; AAB69134.  
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
XX useful in protective vaccines and for diagnosis  
XX Claim 1; Fig 3A-W; 247pp; English.  
XX The present invention describes an isolated and purified nucleic acid (I)  
XX that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.  
XX The 200 kDa outer membrane protein (II) has antibacterial activity and  
XX can be used in vaccines. (II), and its truncated versions, are used as  
XX immunogenic compositions and vaccines to protect against M. catarrhalis  
XX infections, particularly otitis media in humans. (II) is also used as  
XX antigen in immunoassays for detecting specific antibodies (Ab), and to  
XX generate Ab. (I) are used for recombinant production of (II) and its  
XX fragments are used as probes for identifying/cloning 200 kDa protein  
XX genes from other strains, and for diagnostic detection of M. catarrhalis.  
XX (I) makes possible production of large amount of recombinant immunogens.  
XX Expression of truncated versions of (II) reduces toxicity of the protein  
XX towards the Escherichia coli host. The present sequence represents the



QY 1882 gacgggtgcctacaa-----agcttctctattcgtgatgaaaggttcaggaa 1929  
Db 2035 gttgtaccacaaaagatgtagcttaccttggcttagcgaagtagcgtctgacc 2094  
QY 1930 tttaagattagtaactgtattcctaagttaataaccaccaataaccttgagaccatcac 1989  
Db 2095 attggcaaaagacccctaaacacagatggtcttgactgttaagagataccacgaacaaatc 2154  
QY 1990 ttgcagggtgaaacggctacagatcagcaatgacatagccaaaggttaaaatcctaaag-- 2047  
Db 2155 caagtgcgtgtaagttaataattactaatgtgaatggtagtaatccaggtagctgc 2214  
QY 2048 -ttggtattgacccaataaatggttccaccagcctaaagctgacccgtgggtgagcagataaa 2106  
Db 2215 attgcaaataccgctgcctaccagagagataaaattggcttggcttctgatggtgca 2274  
QY 2107 gatggtaaaactcaattggttattgagcaag----- 2137  
Db 2275 gttgatacaaaacacaccttatcttgataagcagaagctacaagttggcaattttaagatt 2334  
QY 2138 ----tggctagcgggttaacgacacacaaacacatcattagaggattgtccccaacactgct 2193  
Db 2335 accaacactggcatttaacgcagggtggttaagccatcacacagggtgtccccaacactgct 2394  
QY 2194 agcattaccatgctggtgctagcagcaccacagaaacagggcaataacataccagcgac 2253  
Db 2395 agcattggcgtatcaagtagc---cgcaacatagaaactgggcaatacaatacca---agac 2448  
QY 2254 gaagcaaatccaaagcccgagtagtcggtgatattataataacaggttttaacctataaa 2313  
Db 2449 aaagcaaatccaaagctccagcattaatgatattataataacaggttttaacctataaa 2508  
QY 2314 aataatagcaactcbggttggcttctctccacttataacacactgttgacttttctgatggc 2373  
Db 2509 aataataacacacccattgacttctctccacttataacacactgttgacttttctgatggc 2568  
QY 2374 aatgcacacacgcgttaagtaactacgatgaacacaaatcaaacacagtaaaacttat 2433  
Db 2569 aatgcacacacgcgttaagtaactacgatgaacacaaatcaaacacagtaaaacttat 2628  
QY 2434 gatgtcaattgtgtagagaaaaccattgaactcacaggcgataatggcaagacaaacaaa 2493  
Db 2629 gatgtgaattgtgtagatacaacacattctataacaggcactgtagacaa---taaaaaa 2685  
QY 2494 attggcgtcaaaacacacacactgaccacaaacaaatgctaatggttaagcaacc----- 2547  
Db 2686 ctggcgtcaaaacacacacactgacaaacaaagtgctaatgtgtaacagcaactaac 2745  
QY 2548 ----aacttttagtacacogataacgatgcctgtttaacgccaagacatcccgaaat 2604  
Db 2746 tttaattgttaactcttagtgatgaagatgccttgtttaacgccaagacatcccgaaat 2805  
QY 2605 ctaaacacctagcgaaggaaattcacaccacaaaggcacagcagacacccctacaa 2664  
Db 2806 ctaaacacctagcgaaggaaattcacaccacaaaggcacagcagacacccctacaa 2865  
QY 2665 acctttaaagtcaaaaaagacggtgca-----actgtacgaaacacatacccggtgggt 2718  
Db 2866 acctttacggttcaaaaagtagatgaataataatgctgtgacgccaacacccatcac 2925  
QY 2719 aaagattgttacacaaaacggcgaacgcttcaacactcttaaaactcaaaaggtgaaacggt 2778  
Db 2926 gtgggtcaaaaagacgcaaaataatcaagtcaaacaccttaaacactcaaaaggtgaaacggt 2985  
QY 2779 cttaacggttctaccataaataagatgtaggttacctttggtcattaaacacccaaagcgt 2838  
Db 2986 cttaattataaacccgacaaaattgtaggttaacctttggcattataacacccaaagcgt 3045  
QY 2839 cttaagccgagcagacacacacttaacaaagatggcttctctattataaacacccgct 2898  
Db 3046 cttaagccgagcagacacacacttaacaaagatggcttctctattataaacacccgct 3102  
QY 2899 agtaacgaacaaatccaaagtcggtgctgtaggtgctgaggttgccaaggt---tgataag 2955

Db 3103 gtagcgaacaaatccaaagtcggtgtagtgcgtgaagtttgccaagtttaataataat 3162  
QY 2956 ggtaatctaaagcacctggcatttgatggcacaacccgctatcaccaaaagatcaaatgtgcttt 3015  
Db 3163 ggtgtgtgtagtgcgtggcatttgatggcacaacccgctattaccagagatgaattggcttt 3222  
QY 3016 actggggttaattggctcacttgataccacaaacccccacactaaccaaaagacagcttaaa 3075  
Db 3223 actgggacttaattggctcacttgataaaaacaaacccccacactaaagcaagac----- 3273  
QY 3076 ggggtgaagtgtgaattaccacaaactggcattaaacgcaaggtggttaaaaagattaccac 3135  
Db 3274 -----ggcattaaacgaggtggttaaaaagattaccac 3306  
QY 3136 attcaatcaggtgtattaccacaaacagcaaatgatcgtgtgacagcggtcggttttat 3195  
Db 3307 attcaatcaggtgtagattggcccaaaacagcctgatcgtgtgacagcggtcggttttat 3366  
QY 3196 gatttaaaacccgaacttgaaagcaaaatcaaacagtgctgttaaaacagacacaaactca 3255  
Db 3367 gatttaaaacccgaacttgaaacaaatcaagcagtaactgcacaaacagacacaaactca 3426  
QY 3256 ttacacgaattctcgtagcagatgaacaaagtaacttaactttacggttagtaaccttac 3315  
Db 3427 ttacacgaattctcgtagcagatgaacaaagtaacttaactttacggttagtaaccttac 3486  
QY 3316 tccagttatgacacctcaaaagcctctgatgtcatcacctttgcaggtgaaacggcatt 3375  
Db 3487 tccagttatgacacctcaaaagcctctgatgtcatcacctttgcaggtgaaacggcatt 3546  
QY 3376 accacaaaggttaaaataaggtgtggtgctgctgagcattgcacaaacaaagccttaacc 3435  
Db 3547 accacaaaggttaaaataaggtgtggtgctgctgagcattgcacaaacaaagccttaacc 3606  
QY 3436 agccttaagctgacccgtgggttaataatggcaaggtgctgctgacagccttaacc 3495  
Db 3607 agccttaagctgacccgtgggttaataatggcaaggtgctgctgacagccttaacc 3666  
QY 3496 ggtcaaaataccatcacagggacttaagcaacacttagcttaattgttaccatga-----t 3549  
Db 3667 ggtcaaaataccatcacagggacttaagcaacacttagcttaattgttaccatgaataaggt 3726  
QY 3550 ggtgaggacacgcactaaagccaaggttgcctaatgacacacacacacacccgctgccgc 3609  
Db 3727 agcgtacgcacacagcaagggcaataataatcaaaagcgaagacacacacccgctgccgc 3786  
QY 3610 agcattgtgtgctgctaaacagcaggtttaaacttgcaagcgaatggtgaaagcgttgac 3669  
Db 3787 agcattgtgtgctgctaaagcagcgtttaaacttgcaagcgaatggtgaaagcgttgac 3846  
QY 3670 ttgtctccacttatgacactgttgactttatcgatggcaatgcccacacccgctaaaggtg 3729  
Db 3847 ttgtctccacttatgacacgcgttcaacttggcgtgcaatggcaccacccgctaaaggtg 3906  
QY 3730 acctatgatgacacaaagcaaaacacagtaaaagtggtctctatgactgataatgtgataataa 3789  
Db 3907 acctatgatgacacaaagcaaaacacagtaaaagtggtctctatgactgataatgtgataata 3966  
QY 3790 acctatgaagtgtacaaagtgataaaaacttggcgtcaaaaacacacacacacacacacacac 3849  
Db 3967 acctatgaagttaaa---gataaaaacttggcgttaaaaacacacacacacacacacacacac 4023  
QY 3850 agtgttaattgttaatgca---accataatttagtgcgcgagtggtgagatgcccttgttaaa 3906  
Db 4024 ggcacaggtgtcaataaaatttgccttaagcaatacgaactactggtgctgctgtgtgcaag 4083  
QY 3907 gccagtgatctgcaccacatcaataaccttggctggcgacatccaaacccgcaaaaggtg 3966  
Db 4084 gccagtgatctgtgtctatcaaaccttctatctggcgacatccaaacccgcaaaaggtg 4143  
QY 3967 gcaagccaaagcagcagctcagcaagctatgtggatgctgtgagcaacaggtcatctat 4026

Db 4144 gcaagccaagcgaacaactcagcagggtatgtggatgctgatggcaataagggtcatctat 4203  
Qy 4027 gacagtaccgataagaagtactatcaagtcaatgacaagggtcaagtggacaaaacaaa 4086  
Db 4204 gacagtaccgataaacaagtactatcaagccaaaatgatggcacagttgataaaaacaaa 4263  
Qy 4087 gaagtccaaaagacaacactggctgcgcccgaagcccaaacccacagatggcacattggctcaa 4146  
Db 4264 gaagtccaaaagacaacactggctgcgcccgaagcccaaacccacagatggcacattggctcaa 4323  
Qy 4147 atgaatgcaaatcagtcattacaagaagcagaagtaaatgatgccaataaaaagcaaggc 4206  
Db 4324 atgaatgcaaatcagtcattacaagaagcagaagtaaatgatgccaataaaaagcaaggc 4383  
Qy 4207 atcaatgaagacacacgctttatcaagggtctgaaacgcccgaagacacacaaaacc 4266  
Db 4384 atcaatgaagacacacgctttgttaaggagcttgaaaagccgcttctgatacaaaaacc 4443  
Qy 4267 aaaaacccgagtaactgtgggtgatttaaatgccgttccccaaaacacgctgaccttt 4326  
Db 4444 aaaaacccgagtaactgtgggtgatttaaatgccgttccccaaaacacgctgaccttt 4503  
Qy 4327 gcagggtatcagcgacaacggtcaaaaactggcgagactttgaccatcaaaagtggtg 4386  
Db 4504 gcagggtatcagcgacaacggtcaaaaactggcgagactttgaccatcaaaagtggtg 4563  
Qy 4387 caaacagacaccaataagctaaaccgataataacatcggtgtggtagcaggtactgtgac 4446  
Db 4564 caaacagacaccaataagctaaaccgataataacatcggtgtggtagcaggtactgtgac 4623  
Qy 4447 ttaactgtcaaaacttgccaagaactaaaccaatttaacagcgttaatgcaaggtggcaac 4506  
Db 4624 ttaactgtcaaaacttgccaagaactaaaccaatttaacagcgttaatgcaaggtggcaac 4683  
Qy 4507 agaattgataaaaagcatctctttgttagcgcgaacacggtcaagccaaagcaaacacc 4566  
Db 4684 aaaaattgataaaaagcggctgtctttgttagactcaagcgggtcaagccaaagcaaacacc 4743  
Qy 4567 cctgtctaagtgcgaatggctggacactgggtggcgaacacgcatcagtaacatcggtgca 4626  
Db 4744 cctgtctaagtgcgaatggctggacactgggtggcgaacacgcatcagtaacatcggtgca 4719  
Qy 4627 gctgttatgataacgctgcggtgaactttaagcagtttaatgaagtggcgaacacggtc 4686  
Db 4780 ----- 4779  
Qy 4687 aacaactaaacacacaaagtaactcaggtgcgtcattaccctttgtgtaacgcatgcc 4746  
Db 4780 ----- 4779  
Qy 4747 aatggcaagcccatcaatggcaccgtaggcaagcccccaaaaagccatcaaggcgccgat 4806  
Db 4780 ----- 4779  
Qy 4807 ggtataactatcacgccaacgccaacggtacactgtggaacaaagatggcaagcccatc 4866  
Db 4780 ----- 4779  
Qy 4867 accgatcggaacaaacttgccaaactgtggcagctcatgtggcaaaccccttgatgaggtcat 4926  
Db 4780 ----- 4779  
Qy 4927 caagtgtggcaagccttagcgccgaactcagatgccaatcaccctcaacacatacaagttcc 4986  
Db 4780 ----- 4779  
Qy 4987 actttgcacaaaattgacacacacacacagggtaagtccaatgcaagggaagcccaagtt 5046  
Db 4780 ----- 4779  
Qy 5047 ctgcccagcctatcagcagcagcaagtaagtgtgctcaggtgtcaaaagtgtgctaaat 5106  
Db 4780 ----- 4779

Qy 5107 gtaggctttaacttgcagaccaatcacaaatcaagtggactttgtcaaaagcctatgatacc 5166  
Db 4780 ----- 4779  
Qy 5167 gtcaactttgtcaatggtacaggtgcgagacatcacaaagcgtgcgtagtgcgtgcagcag 5226  
Db 4780 ----- 4779  
Qy 5227 atgagtaacatcacccgtcaaacaccgcttagcagcgaccgatgatggcgaatgtgctt 5286  
Db 4780 ----- 4779  
Qy 5287 atcaaaagccaagatgtaagtcttcaaaagcagcagcctcatgccaacggtcacta 5346  
Db 4780 ----- 4779  
Qy 5347 aaagcagcgaatacagccagtgatgccaaaactccaactggctaaagccttgttaacccc 5406  
Db 4780 ----- 4779  
Qy 5407 aatgctggtaaagcgatcacagcgatgcgtggtcttctaataacttatcaaaagcggtta 5466  
Db 4780 ----- 4779  
Qy 5467 tttaattccaagatggtacaacttaccacagtaagctctgtaggcatcagtatccaa 5526  
Db 4780 ----- 4779  
Qy 5527 ggcaaaataacagcagcatcacccttaagcaaaagatgggtgaatgtaggcggtaaggtc 5586  
Db 4780 ----- 4785  
Qy 5587 atcagcaatgtggttaaaagcacaacaaagacacccgacgctgcgaatgtacaacagttaaac 5646  
Db 4786 atcagtaatgtggcaaaagcacaacaaagataccgacgctgcgaatgtacaacagttaaac 4845  
Qy 5647 gaagtacgcaactgttgggtcttggtaagtgtgtaataacgctgacggaacacag 5706  
Db 4846 gaagtacgcaactgttgggtcttggtaagtgtgtaataacgctgacggaacacag 4905  
Qy 5707 gtaaacattgccgacatcaaaaaagacccaaattcaggttcaatataacacgcaactgtc 5766  
Db 4906 gtaaacattgccgacatcaaaaaagacccaaattcaggttcaatataacacgcaactgtc 4965  
Qy 5767 atcaaacgagcagcgttacttggcgttaaaagtaataacgataccgaaaacttgcacact 5826  
Db 4966 atcaaacgagcagcgttacttggcgtttaaaggttaataacgataccgaaaacttgcacact 5025  
Qy 5827 ggtgggtataaagtggcgttggtataaagcagcgaacgctaacggtttaaagcaatgtt 5886  
Db 5026 ggtgggtataaagtggcgttggtataaagcagcgaacgctaacggtttaaagcaatgtt 5085  
Qy 5887 tgggtcaaaaaccccaaaaagatggcagcaaaaacccctgtcgcacattataacgcaacgca 5946  
Db 5086 tgggtcaaaaaccccaaaaagatggcagcaaaaacccctgtcgcacattataacgcaacgca 5145  
Qy 5947 ggtcagacccaactatttgaccaacaaccccgagaagcattgacagaataaatagaacaa 6006  
Db 5146 ggtcagacccaactatttgaccaacaaccccgagaagcattgacagaataaatagaacaa 5205  
Qy 6007 ggtatccgcttcttccatgtcaacgattggaacagcctgtggtaacagggcgtaac 6066  
Db 5206 ggtatccgcttcttccatgtcaacgattggaacagcctgtggtaacagggcgtaac 5265  
Qy 6067 ggcattgactcaaatgctcaggaacacactcagtgccgataagtttccagggccaagga 6126  
Db 5266 ggcattgactcaaatgctcaggaacacactcagtgccgataagtttccagggccaagga 5325  
Qy 6127 gatgtaagcgcgctgtgccaatagggcagcaaaaaccccaagcaggaacccaatccatcgcc 6186  
Db 5326 gatgtaagcgcgctgtgccaatagggcagcaaaaaccccaagcaggaacccaatccatcgcc 5385



Db 1019 catttgcttgatcagcatgtaactcctaataacatccgaaaggttactctgtgattaacgatcttt 1078  
Qy 523 attcaaaaccatacagatttacgccaaatacagactcaaatggttctca---gaaatat 579  
Db 1079 attaaaggccatgagattataaagaataacgaagctcaaaagataatgattaaaatat 1138  
Qy 580 agacgcacagcagcagaagacacgcccagctactgactgagtgagccatgcatatgcaaaag 639  
Db 1139 agacgcacaacccgaagcagcagccagctactgagtgagtgagccatgctcatgcaag 1198  
Qy 640 ggtcattttccaaacgctttgttacacggtcaacagctgaagcgaactattctcttgcca 699  
Db 1199 ggtcattttccaaacgctttgttacacggtcaacagctgaagcgaactattctcttgcca 1258  
Qy 700 gtagggttaccgcccgaagccgaaaggatatacatacgtctattggtttctaatagcacaa 759  
Db 1259 gtgggtcttgcgccacagcagggccaaactacatacgtctattggtttctgagcaaca 1318  
Qy 760 gctatcaattatgagcacactagcccttggctgagatacactgagttggttgattacggt 819  
Db 1319 tctagctcgttgagagcgaatagcccttggctgaggtactcgtcagctacagggcaggt 1378  
Qy 820 attgcccctaggttatggttctcagactcttaataataataataata---ataataaa 876  
Db 1379 attgcccctaggttaaggttctgttgcactcagagtgatataataattctagaccggcctat 1438  
Qy 877 gctatgtaccagaaggtaatgggtcaaacataaaaactcgtctaaagccacccggaatgggt 936  
Db 1439 acacaaatacccgagcactagaccccaagtttcaagccaccaataatacgaagggcgtt 1498  
Qy 937 ttattttccattggttagcactatcaacgcgttaaaatcatcattcgtcgtgtaggttat 996  
Db 1499 ccactttccattggttagtactctataaacgttaaaatcatcattcgtcgtgtaggttat 1558  
Qy 997 gaggataccgctcgggtccaatgtggcagcgttaaaagcgtgagaaactcggcta---ag 1053  
Db 1559 aataaaaccgctcgggtccaatgtggcagcgttagaagcgttgggttaaggg 1618  
Qy 1054 cgtcaattacttttaagggtgatgataacggttactggcgttaagaaacacactgggcgag 1113  
Db 1619 cgtagaattactttcaggttgatgataacagctactgacgtacgtataaaataggtttgataat 1678  
Qy 1114 actttaaccattaaaggtgtagacccaagcggacaaagcgaactaacctgataataataacatt 1173  
Db 1679 actttaactataaaggtggtg---cagagaccaacgattaacccgataataatcgtt 1735  
Qy 1174 ggttggttaacagataataactggtctgaaagttaaaactgcttaaaacacctaagcgtt 1233  
Db 1736 gtggtaaaagagcgtgataatagtggtcgtgaaagttaaaactgcttaaaacacctaagcgtt 1795  
Qy 1234 ctggaacaggttagcaccacaaacacctaagcgcagcagagagaagttagcgttag---gtagt 1290  
Db 1796 cttagtgagtgatacaactacattataatgccaacaccacagattaaaggttaggttagt 1855  
Qy 1291 ggttaataaccgctgagctacaaagcgttggttttaacctttacc---caacaa 1344  
Db 1856 agtagtactacagctgaattatgagtgatagtttaacctttaccgcccacatacagc 1915  
Qy 1345 aatgcaagcagacaaaaacgctctatggtcactgagcgtgaggttactgataaactct 1404  
Db 1916 agtcaaaagcaagcaaaaacgctctatggtggttaagtgggtgagtttactataatagca 1975  
Qy 1405 aatacggcacttga---gatactactcgtatcaccaagaataaaaattggttttagcaat 1461  
Db 1976 gaaacaacagcagaactcggcactactcgtattaccagagataaaaattggtttgctcga 2035  
Qy 1462 aagcctggtacagttgatgaaacaaacacttatcttgataaaagacagcgttaaaag--- 1516  
Db 2036 gatgggtgatgttgataaaacagcaccattatttgataaaaaaacacacttaagtggtt 2095  
Qy 1517 -----tt 1518  
Db 2096 agtgttgcaattaccatagacaatggcattgtagcagggtataaaaaagatcagtaactctt 2155

Qy 1519 ggcaacagcagccctaaacaacggttggtgactgtttaataacaccattggtggtagcaat 1578  
Db 2156 gccaaaggttagcagtgtataacgtagcgtttaccatcgaacagctcaaaagccgcaagcct 2215  
Qy 1579 aaacaaatccaagtgcgtgctgagtcgattaaatttcccgatgtgaattgtaattgatatca 1638  
Db 2216 actttaacagcagcgtgcgtcagtgctcacacctactgaaatcatcagttgtagtcaag 2275  
Qy 1639 aatggcccaaaatcggcactactcgtattacggaagagaaatttgctttgctgtagct 1698  
Db 2276 agtggcaattgttaccggcccaacttacaacattggcgtgaaacaccagcgtttaacagt 2335  
Qy 1699 gatggttaaaagtgtataa----- 1715  
Db 2336 gatggcactagtataaatttagtgttaagggttagtggtacgaaacatagcttagttacc 2395  
Qy 1716 -----aaagtcaccattatttgataaaaaaacacttcaagtggtgggtggtt 1761  
Db 2396 gccgaacatttggcagactatcaaatggaagtcaatcgaaacggtcagctgactctcaaa 2455  
Qy 1762 aaaaattaccaaaagacagtgccattaatgcaggtgatcaaaagatcagtaattgttaaaagt 1821  
Db 2456 agctttaaccgttaaaagaagacagatgtagcgaacacgctatcaacggtggttaaaagt 2515  
Qy 1822 gcaacgacatgataccgctcagtcacttataaacagcttaaacaaagtcaccaagaagcgc 1881  
Db 2516 acgacaaaaaatgcggcgagtcagctcttaaaactcaaaaggtataaaacgggtctaagc 2575  
Qy 1882 gacggtgcccctacaa-----agcttctctattcgtgatgaaaaaggttcagga 1929  
Db 2576 gttgctaccaaaaagatggtacggttacccttgggttagcgaagtagcgttctgacc 2635  
Qy 1930 ttacagattagtaactgtattcttaattggttaataaccacaaatcctttgagaccatcacc 1989  
Db 2636 attggcaaaagcaccctaaacaacagatgggttgaactgttaaaagtatacaacgaacaaatc 2695  
Qy 1990 ttgycaggtgaaacggcgtacgtatcagcaatgacatagcgaaggttaaaaggtcaaaag-- 2047  
Db 2696 caagtcggtgtaagtgcattaaatttactaatgtgaatgtagtaataccaggctactggc 2755  
Qy 2048 -ttggtattgacccaatacaatggtctcaccacgctgaagtcacggtgggttagcgataaa 2106  
Db 2756 attgcaataccgctcgtcattaccagagataaaaattggttctggttctgtagtggtgca 2815  
Qy 2107 gatgtaaaaactcaattggtttattgagcaag----- 2137  
Db 2816 gttgatacaaaaacacttatcttgatcaagacaagctacaagttggcaattgttaagatt 2875  
Qy 2138 ----tggtgaggtgaacgacacacacacacacatcatttagaggttgtccccaacactgctt 2193  
Db 2876 accaacactggcattaaacgcaggtggtaagccatcacagggtgtccccaacactgctt 2935  
Qy 2194 agcattaccaatgaggtggtgtagcaccacacagacagggcaatacaataccacagcagac 2253  
Db 2936 agcattccgatcaaaatagc---cgcaacatagaactgggcaatacaatcca---agac 2989  
Qy 2254 gaagacaaaatccaaagccgcagctatcgtgatatataataacagcgttttaacctaaaa 2313  
Db 2990 aaagacaaaatccaaacgctgccagcattaaatgatatataatacagcgttttaacctaaaa 3049  
Qy 2314 aataatagcaactcgttgggttcttctccacttataacactttagctttttagctttagctggc 2373  
Db 3050 aataatacaacccactgacttcttctccacttatgacattgtgacttttggccaatggc 3109  
Qy 2374 aatgcaaccccgctaaaggttaactttagcgtgaaacccaatacaacacagtaaaagttaactat 2433  
Db 3110 aatgcaaccccgctaaacagtaaacccatgataccgctacacaaacacagtaaaagtgttat 3169  
Qy 2434 gatgtcaatgtgatgaaaaacacattgaaactcacagcgtgataatggcaagacacacaa 2493  
Db 3170 gatgtgaatgtgatgatacaaacatttcatcctaacagcgtgactgtagacaa---taaaaa 3226



Qy	2494	a ttg c g c g t c a a a a c c a c c a c t g a c c a c a n c a a a t g c t a a t g g t a a a g c a a c c - - - - -	2547
D b	3227	c t t g c g t c a a a a c c a c c a a c t g a c a a a a c a a g t g t a a t a c a g c a a c t a a c	3286
Q y	2548	- - - a a c t t t a g t a c c o g a t a a c g a t g c c t t g t t a a c g c c a a g a g a t g c c g a a a t	2604
D b	3287	t t a a t g t t a a c t c t a g t a l g a a g a t g c c t t g t t a a c g c c a a g a g a t g c c g a a a t	3346
Q y	2605	c t a a c a c c t a g c a a g a a a t t c a c a c c a a a g g c a c a g c a c a g c c c t a c a a	2664
D b	3347	c t a a a c c c t g c a a g g a a t t c a c a c c a a a a g g c a c a g c a c a g c c c c t a c a a	3406
Q y	2665	a c c t t t a a a g t c a a a a a g c g g t g c a - - - - - a c t g a t g a c g a a a c c a c c g t g g g t	2718
D b	3407	a c c t t t a c c g t t a a a a g t a g a t g a a a t a a t a a t g c t g a t g a c g c c a c c a c c a c	3466
Q y	2719	a a a g a t g g t a c a a a a c g c a a g c g t c a a c t c t a a a a c t a a a a g g t g a a a c g g t	2778
D b	3467	g t g g g t c a a a a g a c g a a t a a t a a g t c a a c c c t a a c t a a a a g g t g a a a a c g g t	3526
Q y	2779	c t a a c g g t g t c t a c c a a t a a g a t g t a c g g t t a c c t t t g g c a t t a a c c c c a a a g c g g t	2838
D b	3527	c t t a a t t a a a a c g a c a a a a t g g t a o g g t t a c c t t t g g c a t t a a c c a c c a a g c g g t	3586
Q y	2839	c t t a a a g c g g a c a g a c c a c t c t a a a c a a a g a t g g c t t g t c t a t t a a a a c c c c g c t	2898
D b	3587	c t t a a a g c g g a a a a g c a c c - - - c t a a a c g a c g g t g g g t t g t c t a t t a a a a c c c a c t	3643
Q y	2899	a g t a a c g a a a a t c c a a g t c g g t g c t g a t g g c g t g a a g t t t g c c a a g g t - - - t g a t a a g	2955
D b	3644	g g t a g c a a a a t c c a a g t c g g t g c t g a t g g c g t g a a g t t t g c c a a g g t t a a t a a t a t	3703
Q y	2956	g g t a a t c a a g c a c t g c a t t g a t g a c a a g c g t a t c a c a a a g a t c a a a t t g g c t t t	3015
D b	3704	g g t g t t a g t g c t g a t g t g g c a c a c t g c a t t a c c a g a g a t g a a t t g g c t t t	3763
Q y	3016	a c t g g g g t a a t g c t c a c t t a t a c c a c a a a c c c a c c a c t a a c c a a a g a c a a g c t t a a a	3075
D b	3764	a c t g g a c t a a t g c t c a c t t g a t a a a a g c a a a c c c a c c a c t a a g c a a a g a c - - - - -	3814
Q y	3076	g t g g g t g a a g t g a a a t t a c c a a c a c t g g a t t a a c g a g t g g t t a a a a g a t t a c c a a c	3135
D b	3815	- - - - - g g c a t t a a c g a g g t g g t a a a a g a t t a c c a a c	3847
Q y	3136	a t t c a a t c a g t g a t a t t a c c a a a a c a g c a a t g a t g c t g a c a g g c g g t c g g g t t a t	3195
D b	3848	a t t c a a t c a g t g a g a t t g c c a a a c g e c c a t g a t g c t g t g a c a g g c g g c a a g a t t a t	3907
Q y	3196	g a t t a a a a a c c g a a c t t g a a g a a a a t c a a c a g t g c t g t a a a c a g a c a a a c t c a	3255
D b	3908	g a t t a a a a c c g a a c t t g a a a c a a a t c g c g a g t a c t g c c a a a c a g c a a a c t c a	3967
Q y	3256	t t a c a g a a t t c a g t a g c a g a t g a a c a a g g t a a t c a c t t t a c g g t t a g t a a c c c t t a c	3315
D b	3968	t t a c a g a a t t c t c a g t a g a t g a a c a a g t a a a t a c t t t a c g g t t a g t a a c c c t t a c	4027
Q y	3316	t c c a g t a t g a o c c t a a a a g c c t c t g a t g t c a t c a c c t t t g a g g t g a a a c g c a t t	3375
D b	4028	t c c a g t t a g a c c c t a a a a c c t g a t g t c a t c a c c t t t g a g g t g a a a c g g c a t t	4087
Q y	3376	a c c a c a a g g t a a a t a a a g g t g t g g t g c g t g g g c a t t g a c c a a c c a a a g g c t t a a c c	3435
D b	4088	a c c a c a a g g t a a a t a a a g g t g t g t g c g t g g g c a t t g a c c a a c c a a a g g c t t a a c c	4147
Q y	3436	a c g c c t a a g c t g a c c g t g g g t a a t a a t g g c a a a g g c a t t g t c a t t g a c a g t a a a g a t	3495
D b	4148	a c g c c t a a g c t g a c c g t g g g t a a t a a t g g c a a a g g c a t t g t c a t t g a c a g c c a a a t	4207
Q y	3496	g g t c a a a t t a c c a t c a c a g g a c t a a g c a c c t c t a g c t a a t g t t a c c a a t g a - - - - -	3549
D b	4208	g g t c a a a a t t a c c a t c a c a g a c t a a g c a c a c t c t a g c t a a t g t t a c c a a t g a t a a g g t	4267
Q y	3550	g g t g a g g a c a c g c a c t a a g c c a a g g g c t t g c a a t g a c a c g a c a a a a c c c g t c c g c c	3609



Db 5321 ----- 5320  
QY 4687 acaacctaacaacaaagtaactcaggtgctgctattaccctttgtggtaacccgatgcc 4746  
Db 5321 ----- 5320  
QY 4747 aatggcaagcccatcaatggcacccgatggcaagcccccaaaaagcccatcaagggcgccgat 4806  
Db 5321 ----- 5320  
QY 4807 ggtataactatcacgccaacgccaacgcttacctgtggcaaaagatggcaagcccatc 4866  
Db 5321 ----- 5320  
QY 4867 accgatggacaaaacttgccaatctggcagctcatggcaaaccccttgatgcaggtcat 4926  
Db 5321 ----- 5320  
QY 4927 caagtggtggaagccttagggcggaactcagatggcatcaccctaaaccaacatcaagttcc 4986  
Db 5321 ----- 5320  
QY 4987 acttggccaaaattgacacacacacacaggttaatgccaaatgcaggggcaagcccaagt 5046  
Db 5321 ----- 5320  
QY 5047 ctgcccagcctatcagcagcacagcaaaagttaatgctgcccagtgtaaaagtgtgctaaat 5106  
Db 5321 ----- 5320  
QY 5107 gtaggctttaactgacagaccaatcaaatcaagtggaactttgtcaaaagcctatgatacc 5166  
Db 5321 ----- 5320  
QY 5167 gtcaactttgcaatggtacaggtgcgcacatcacaaagctgcgtagtgcgtatggcagc 5226  
Db 5321 ----- 5320  
QY 5227 atgagtaacatcacctgtaacacacgccttagcagcgaccgatgatgatggcaatgtgctt 5286  
Db 5321 ----- 5320  
QY 5287 atcaaaagcaaaagatggtaagttcttacaagcagacacatcagccaacagcgctacta 5346  
Db 5321 ----- 5320  
QY 5347 aaagcaggcaaatcagccagtgatgccaaaactccaactggttctaagccttgttaacccc 5406  
Db 5321 ----- 5320  
QY 5407 aatgctggtaaaggcagtagcagcgatgcagtgagtgcttctaataacttatcaaaagcgta 5466  
Db 5321 ----- 5320  
QY 5467 tttaaatccaaagatggtacaactactaaccacagtaagctctgatggcatcagtatccaa 5526  
Db 5321 ----- 5320  
QY 5527 ggcaaaagataacagcagcatcaccccaagcaagatgggctgaatgtaggcggttaaggtc 5586  
Db 5321 ----- 5326  
QY 5587 atcagcaatggtggttaagcacaacaaagacacccgacgctgccaatgtacacagttaaac 5646  
Db 5327 atcagtaatggtggcaaaagcacaagataccgacgctgccaatgtacacagttaaac 5386  
QY 5647 gaagtaacgaacttgttgggtcttggtaatgctggtaatgatcaacgctgcagcgaatcag 5706  
Db 5387 gaagtaacgaacttgttgggtcttggtaatgctggtaatgatcaacgctgcagcgaatcag 5446  
QY 5707 gtaaacattgccacatcaaaaagacccaaattcaggttcatcatcattaccgacactgtc 5766  
Db 5447 gtaaacattgccgacatcaaaaagacccaaattcaggttcatcatcattaccgacactgtc 5506

QY 5767 atcaaaagcagcaggttacttggcggtaaaaggttaataacgataccgaaaaacttggccact 5826  
Db 5507 atcaaaagcagcaggttacttggcggtaaaaggttaataacgataccgaaaaacttggccact 5566  
QY 5827 ggtgggtgtacaagtggcgtggataaaagcggcaacgctaacgctaacggcgatttaagcaatgtt 5886  
Db 5567 ggtgggtgtacaagtggcgtggataaaagcggcaacgctaacgctaacggcgatttaagcaatgtt 5626  
QY 5887 tgggtcaaaaccccaaaaagatggcagcaaaaagccctgctgcacatttaaacgcgcga 5946  
Db 5627 tgggtcaaaaccccaaaaagatggcagcaaaaagccctgctgcacatttaaacgcgcga 5686  
QY 5947 ggtcagaccactatttgaccaacaaccccgcaagaccatttgacagataaaatgaacaa 6006  
Db 5687 ggtcagaccactatttgaccaacaaccccgcaagaccatttgacagataaaatgaacaa 5746  
QY 6007 ggtatccgcttcttccatgtcaacgatggcaatcaaaagcctgtgtgtacaagggcgtaac 6066  
Db 5747 ggtatccgcttcttccatgtcaacgatggcaatcaaaagcctgtgtgtacaagggcgtaac 5806  
QY 6067 ggcattgactcaagtgcctcagcaagcactcagtgccgataggtttccagggccaagga 6126  
Db 5807 ggcattgactcaagtgcctcagcaagcactcagtgccgataggtttccagggccaagga 5866  
QY 6127 gatggtaagcccgcttgcctataggcagacaacccaagcaggaaccaatccatcgcc 6186  
Db 5867 gatggtaagcccgcttgcctataggcagacaacccaagcaggaaccaatccatcgcc 5926  
QY 6187 atcgggtgataacgcacaagccacggcgatcaatccatccatccatccatccatccatcc 6246  
Db 5927 atcgggtgataacgcacaagccacggcgatcaatccatccatccatccatccatccatcc 5986  
QY 6247 gtaacaggtgaagcactctgtgccatcgcgacccaagcactgttaaggtgtataacagt 6306  
Db 5987 gtaacaggtgaagcactctgtgccatcgcgacccaagcactgttaaggtgtataacagt 6046  
QY 6307 tacagtggtggtaataacaacccagtttatcgatgccactcagacagatgtctttgtgtgtg 6366  
Db 6047 tacagtggtggtaataacaacccagtttatcgatgccactcagacagatgtctttgtgtgtg 6106  
QY 6367 ggaataaatacaccgtgacggaaagttaactcggttgccttaggttcaaaactctgcacatc 6426  
Db 6107 ggaataaatacaccgtgacggaaagttaactcggttgccttaggttcaaaactctgcacatc 6166  
QY 6427 agtgacggcacacgcaggcacacacagccaacaaaatctgcagcgcacagcaggtacaaac 6486  
Db 6167 agtgacggcacacgcaggcacacacagccaacaaaatctgcagcgcacagcaggtacaaac 6226  
QY 6487 accacagcaggtgcaacaggttacggtttaaaggtttgtgtgacaaaacggcggttgggtgcg 6546  
Db 6227 accacagcaggtgcaacaggttacggtttaaaggtttgtgtgacaaaacggcggttgggtgcg 6286  
QY 6547 gttctccgtgggtgctcaggtgctgaacgcgcgtatccaaaatgtggcagcaggtggaagtc 6606  
Db 6287 gttctccgtgggtgctcaggtgctgaacgcgcgtatccaaaatgtggcagcaggtggaagtc 6346  
QY 6607 agtgccaccagcagcaggtgctgaatggtagccagttgtacaagccaccccaagcaggt 6666  
Db 6347 agtgccaccagcagcaggtgctgaatggtagccagttgtacaagccaccccaagcaggt 6406  
QY 6667 gccaaacgaacaaatgagcttgaccatcgatccacaaaacgaacaaataaaagccaatgca 6726  
Db 6407 gccaaacgaacaaatgagcttgaccatcgatccacaaaacgaacaaataaaagccaatgca 6466  
QY 6727 gggatttcatcagcgatggcgatggcgatggcgatggcgatggcgatggcgatggcgatggcg 6786  
Db 6467 gggatttcatcagcgatggcgatggcgatggcgatggcgatggcgatggcgatggcgatggcg 6526  
QY 6787 atggttaccgggggtattgccaacccacaaaggttcaaggtgcgggtggcgatggcgatggcg 6846  
Db 6527 atggttaccgggggtattgccaacccacaaaggttcaaggtgcgggtggcgatggcgatggcg 6586

QY	6847	aagctgcggaataaggtcaatgggtatttaaaatcaatgattcagccgatacccaagggc	6906
Db	6587	aagcgtcggataaaggccaatgggtatttaaaatcaatgattcagccgatacccaagggc	6646
QY	6907	catgtaggggcgccaggttggtgcaggttttcacttt	6942
Db	6647	catgtaggggcgccaggttggtgcaggttttcacttt	6682
RESULT	6		
AAT38740			
ID	AAT38740	standard; DNA; 6971 BP.	
AC	AAT38740;		
XX	25-JAN-1997	(first entry)	
DT			
XX	Moraxella outer membrane protein gene.		
DE			
XX	Outer membrane protein; OMP; Immunogen; vaccine; otitis media;		
KW	diagnosis; ss.		
XX	Mycobacterium catarrhalis strain 4223.		
OS			
XX	Key	Location/Qualifiers	
FH	CDS	706..6684	
FT		/*tag= a	
FT			
XX	WO9634960-A1.		
PN			
XX	07-NOV-1996.		
PD			
XX	29-APR-1996;	96WO-CA00264.	
PF			
XX	26-MAR-1996;	96US-0621944.	
PR	01-MAY-1995;	95US-0431718.	
PR	07-JUN-1995;	95US-0478370.	
XX	(CONV-) CONNAUGHT LAB LTD.		
FA			
XX	Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;		
PI			
XX	WPI; 1996-506162/50.		
DR	P-PSDB; AAW04505.		
XX	Moraxella outer membrane protein - useful as immunogen in protective		
PT	vaccine and for diagnosis		
PT			
XX	Claim 14; Fig 6; 109pp; English.		
PS			
XX	A genomic DNA fragment (T38740) of Moraxella catarrhalis otitis		
CC	media strain 4223 includes the coding region for a 1992-amino		
CC	acid protein (W04505) identified as an approx. 200 kDa outer		
CC	membrane protein (OMP). The DNA was isolated from a strain 4223		
CC	genomic library in phage lambda EMBL3 by screening with an		
CC	anti-200 kDa protein guinea pig antiserum. The gene can be		
CC	used for the recombinant expression of the OMP (for use in		
CC	vaccines), for the prep. of hybridisation probes, or may be		
CC	incorporated into a live vector for use in direct immunisation.		
XX	Sequence 6971 BP; 2265 A; 1553 C; 1533 G; 1620 T; 0 other;		
SQ			
Query Match	37.2%;	Score 2580.6;	DB 17; Length 6971;
Best Local Similarity	66.9%;	Pred. No. 0;	
Matches 4803;	Conservative	0; Mismatches 1104;	Indels 1270; Gaps 29;
QY	1	atgaatcacatctcataaagtcattttaaacaagccacagggcacacatttatggcgtggca	60
Db	540	atgaatcacatctcataaagtcattttaaacaagccacagggcacacatttatggcgtggca	599
QY	61	gagtcgcgcaaaatccacacg-cgagggaggtagcagtagtaccgcagacaggtggcag	119



||||| 3846 catccaatcagtgagattgccccaaacagccatgatctgtgacaggcggaagattta 3905  
||||| 3195 tgatttaaaacccgaaacttgaaagcaaaatcaacagtgctgcttaaaacagcacaaaactc 3254  
||||| 3906 tgatttaaaacccgaaacttgaaacaaaatcagcagtgactgccccaaacagcacaaaactc 3965  
||||| 3255 attacacgaattctcagtagcagatgaacaagtgtaatacactttacggttagtaaccctta 3314  
||||| 3966 attacacgaattctcagtagcagatgaacaagtgtaatacactttacggttagtaaccctta 4025  
||||| 3315 ctccagttatgaacacctcaaaacacctctgatgtcatcactttgcagtgaaaacggcat 3374  
||||| 4026 ctccagttatgaacacctcaaaacacctctgatgtcatcactttgcagtgaaaacggcat 4085  
||||| 3375 taccacaaaggtaaaataaagggtgtggtgcgtggtggcatttgaccacaaacaaaggcttaac 3434  
||||| 4086 taccacaaaggtaaaataaagggtgtggtgcgtggtggcatttgaccacaaacaaaggcttaac 4145  
||||| 3435 cagcgttaagctgacccgtgggttaataatggcaaggcattgtcattgacagtaaga 3494  
||||| 4146 cagcgttaagctgacccgtgggttaataatggcaaggcattgtcattgacagtaaga 4205  
||||| 3495 tggcctaaatcaccatcacaggaactgaagcaacacctctagcttaattgttaccatga 3548  
||||| 4206 tggcctaaatcaccatcacaggaactgaagcaacacctctagcttaattgttaccatga 4265  
||||| 3549 tggcgaggacagccactaagccaaaggcttgccaatgacacagcaaaacccgtgcgc 3608  
||||| 4266 tagcgtacgacacagaaacagggcaataataaacaagacgagcaaaacccgtgcgc 4325  
||||| 3609 cagcattgggtgatgtctaaacgagcgtttaaacttgcaaggcaatgggtgaagcggtga 3668  
||||| 4326 cagcattgggtgatgtctaaacgagcgtttaaacttgcaaggcaatgggtgaagcggtga 4385  
||||| 3669 ctttgtctccacttatgacactgttgacttta togtatggcaatggccacacccgttaagg 3728  
||||| 4386 ctttgtctccacttatgacactgttgacttta togtatggcaatggccacacccgttaagg 4445  
||||| 3729 gacctatgatgaacaaagcaaaacccagtaagtggtctatgatgtcaatgtggataataa 3788  
||||| 4446 gacctatgatgaacaaagcaaaacccagtaagtggtctatgatgtcaatgtggatgatc 4505  
||||| 3789 aaccattgaagtgaacagtgataaaacacttggcgtcaaaacacacacacacacacacac 3848  
||||| 4506 aaccattgaagtgaacagtgataaaacacttggcgtcaaaacacacacacacacacacac 4562  
||||| 3849 aagtgctaattggttaatgaac 3905  
||||| 4563 tggcacaggtgctaataattgcccataagcaatacaagctactggtggtggtgctcaa 4622  
||||| 3906 agccagtgatatgcccacccatcaaaatcaccttggcgtgcaacacacacacacacacacac 3965  
||||| 4623 tggcacaggtgctaataattgcccataagcaatacaagctactggtggtggtgctcaa 4682  
||||| 3966 ggaagcaagcaagcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4025  
||||| 4683 ggaagcaagcaagcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4742  
||||| 4026 tgacagtaccgataagaagtaactcaagctcaatgacaggggtcaagtggcacaacacacac 4085  
||||| 4743 tgacagtaccgataagaagtaactcaagctcaatgacaggggtcaagtggcacaacacacac 4802  
||||| 4086 agaagttgcaaaagcaaaactggtgcccagaagcccaacccacagatggcacttggctca 4145  
||||| 4803 agaagttgcaaaagcaaaactggtgcccagaagcccaacccacagatggcacttggctca 4862  
||||| 4146 aatgaatgtaaatcagtcatttaacaaagcaagtaaatgatgccaataaaagcaagg 4205  
||||| 4863 aatgaatgtaaatcagtcatttaacaaagcaagtaaatgatgccaataaaagcaagg 4922  
||||| 4206 catcaatgaagcaaacgcctttatcaaaagggttgaaaacggcgccaaagacacacacac 4265  
|||||

4923 catcaatgaagcaaacgcctttgttaaaaggacttgaaaaagcgcgtttctgtatacaaaaaac 4982  
4266 caaaaaacgcgcagtaactgtgggtgatttaaatgccttgccccaaaacacccgctgacctt 4325  
4983 caaaaaacgcgcagtaactgtgggtgatttaaatgccttgccccaaaacacccgctgacctt 5042  
4326 tgcaggggatacagggcacacacgcgttaaaaaactggggcagagactttgaccatcaaaagg 4385  
5043 tgcaggggatacagggcacacacgcgttaaaaaactggggcagagactttgaccatcaaaagg 5102  
4386 gcaaacagacacacaaataagctaaacgataaatacctcgtgtggttagcaggttactgatatg 4445  
5103 gcaaacagacacacaaataagctaaacgataaatacctcgtgtggttagcaggttactgatatg 5162  
4446 ctctactgtcaaaacttgccaaagacacacacacacacacacacacacacacacacacacac 4505  
5163 ctctactgtcaaaacttgccaaagacacacacacacacacacacacacacacacacacacac 5222  
4506 cagaatttgataaaaaagcctctctttttagacgcgaacacgcgtcaagcccaaaagcaaacac 4565  
5223 caaaatttgataaaaaagcgtgctctttagacacacacacacacacacacacacacacacac 5282  
4566 ccttgtctaaagtgcac 4625  
5283 ccttgtctaaagtgcac 5319  
4626 agctgtgtgataacgcgtgcgttgaaactttaagcagctttaagcagctttaagcagcttta 4685  
5320 ----- 5319  
4686 caacaacctaaac 4745  
5320 ----- 5319  
4746 caatggcaagcccatcaatggcaacgatggcaacacacacacacacacacacacacacacac 4805  
5320 ----- 5319  
4806 tggtaatactatcacgcac 4865  
5320 ----- 5319  
4866 caccgatgggac 4925  
5320 ----- 5319  
4926 tcaagtgggtgcaagcctagggcgcaactcagatgacctcacccctaacacacacacacacac 4985  
5320 ----- 5319  
4986 cactttgcac 5045  
5320 ----- 5319  
5046 tctgcccagcctatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5105  
5320 ----- 5319  
5106 tgtaggctttaactgtcagac 5165  
5320 ----- 5319  
5166 cgtcaactttgtcaatgggtacaggtgcccagacacacacacacacacacacacacacacacac 5225  
5320 ----- 5319  
5226 gatgagtaacatcacccgtcaac 5285  
5320 ----- 5319  
5286 tatcaaaagcaaaagatgggttaagtctctcaaaagcagacgacacacacacacacacacacac 5345  
5320 ----- 5319

QY 5346 aaagcaggcaaatcagccagtgatgcacaaactccaactggtctaagccttgtaaccc 5405  
Db 5320 ----- 5319  
QY 5406 caatgctggttaagcagctacagcgatgcagtggtctcttaataacttatcaaaagcgt 5465  
Db 5320 ----- 5319  
QY 5466 atttaaatccaaagatggtataactactaccacagtaagctctgatggcatcagatcca 5525  
Db 5320 ----- 5319  
QY 5526 aggcagaagataacagcagcatcacccctaagcaaaagatggtcgtgaatgtagcggtaaggt 5585  
Db 5320 -----aaggt 5324  
QY 5586 catcagcaatggtgtaaaagcacaagaacacagcagctgccaatgtacaacagttaaa 5645  
Db 5325 catcagtaagtgtggcacaagcacaagaatcacgcagctgccaatgtacaacagttaaa 5684  
QY 5646 cgaagtcgcaactgttggtctcttggttaattgctggttaataagataacgctgacggcaatca 5705  
Db 5385 cgaagtcgcaactgttggtctcttggttaattgctggttaataagataacgctgacggcaatca 5444  
QY 5706 ggttaacattgcacatcaaaaaaagcccaaaattccaggttcacatctaacgcactgt 5765  
Db 5445 ggttaacattgcacatcaaaaaaagcccaaaattccaggttcacatctaacgcactgt 5504  
QY 5766 catcaagcagcagcagctacttgcgttaagtaataagcattacaagcagcaaaactggcac 5825  
Db 5505 catcaagcagcagcagcagctacttgcgttaagtaataagcattacaagcagcaaaactggcac 5564  
QY 5826 tgggtgtacaagtggcgtggataaagacggcaacgctaaacggcagatttaagcaatgt 5885  
Db 5565 tgggtgtacaagtggcgtggataaagacggcaacgctaaacggcagatttaagcaatgt 5624  
QY 5886 ttgggttaaaacccaaagataggcagcaaaaaagccctcgtcgcacttataacgcgc 5945  
Db 5625 ttgggttaaaacccaaagataggcagcaaaaaagccctcgtcgcacttataacgcgc 5684  
QY 5946 aggtcagaccaaactatttgaccacaaccccgagagagccattgacagataaatagaaca 6005  
Db 5685 aggtcagaccaaactatttgaccacaaccccgagagagccattgacagataaatagaaca 5744  
QY 6006 aggtatccgtcttccatgctcaacagatggcaatcaagagcctgtgtgtaagggcgtaa 6065  
Db 5745 aggtatccgtcttccatgctcaacagatggcaatcaagagcctgtgtgtaagggcgtaa 5804  
QY 6066 cggcattgactcaagtgcctcaaggcagcactcagtggtgagataggtttccagggccaagtc 6125  
Db 5805 cggcattgactcaagtgcctcaaggcagcactcagtggtgagataggtttccagggccaagtc 5864  
QY 6126 agatggtgaagcccgcttgccatagcagcaacaacccaagcagcagcaaatccatcgc 6185  
Db 5865 agatggtgaagcccgcttgccatagcagcaacaacccaagcagcagcaaatccatcgc 5924  
QY 6186 catcgggtgaatacgcacaaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6245  
Db 5925 catcgggtgaatacgcacaaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5984  
QY 6246 ggttaacaggtaagcactctggtgcacatcgcgacccaagcactgttgaagcgtgataacag 6305  
Db 5985 ggttagcaggtaagcactctggtgcacatcgcgacccaagcactgttgaagcgtgataacag 6044  
QY 6306 ttacagtgtgggttaataacacagcttatacgatgacctcagcagcagcagcagcagcagcagc 6365  
Db 6045 ttacagtgtgggttaataacacagcttatacgatgacctcagcagcagcagcagcagcagcagc 6104  
QY 6366 gggcaataacatcacctgacccgaaagtaactcgtgtgctttaggttcaaaactcgcact 6425  
Db 6105 gggcaataacatcacctgacccgaaagtaactcgtgtgctttaggttcaaaactcgcact 6164

QY 6426 cagtgcaggcacacacagcagcgacacaaagccaaataatctgacggcgacagcgatacaac 6485  
Db 6165 cagtgcaggcacacacagcagcgacacaaagccaaataatctgacggcgacagcgatacaac 6224  
QY 6486 caccacagcaggtgcaacagctacaggttaaaagccttgctgggacaaacgagcgggtgctgc 6545  
Db 6225 caccacagcaggtgcaacagcctacaggttaaaagccttgctgggacaaacgagcgggtgctgc 6284  
QY 6546 ggttcctcgtgggtgctcctcaggtgctgaacccgctatccaaatgtgagcagcaggtgaggt 6605  
Db 6285 ggttcctcgtgggtgctcctcaggtgctgaacccgctatccaaatgtgagcagcaggtgaggt 6344  
QY 6606 cagtgcacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6665  
Db 6345 cagtgcacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6404  
QY 6666 tgccaaacgcacaaatgagcttgaccatgctgacccatgctatcccaaaacgaaataaagccaatgc 6725  
Db 6405 tgccaaacgcacaaatgagcttgaccatgctgacccatgctatcccaaaacgaaataaagccaatgc 6464  
QY 6726 agggatttcacagcagatggcgatggcgctccatgcccaagcctacattccttgccagatc 6785  
Db 6465 agggatttcacagcagatggcgatggcgctccatgcccaagcctacattccttgccagatc 6524  
QY 6786 catggttacccgggggtattgccaccacacacgggtcgaaggtgcgtgagcagtgagcagtc 6845  
Db 6525 catggttacccgggggtattgccaccacacacgggtcgaaggtgcgtgagcagtcagtc 6584  
QY 6846 gaagctcgcagataatggtcaatggttatttaaatacaatggttcagccagcagcagcagcagc 6905  
Db 6585 gaagctcgcagataatggtcaatggttatttaaatacaatggttcagccagcagcagcagcagc 6644  
QY 6906 ccattgtagggcgcgagctggtgcaggtttttcacttt 6942  
Db 6645 ccattgtagggcgcgagctggtgcaggtttttcacttt 6681

RESULT 7  
AAF59100  
ID AAF59100 standard; DNA: 6973 BP.  
XX  
AC AAF59100;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa gene SEQ ID NO:1.  
XX  
KW Moraxella catarrhalis strain 4223; major outer membrane protein;  
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection; ds.  
XX  
OS Moraxella catarrhalis.  
XX  
PN WO200107619-A1.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-CA00870.  
XX  
PR 27-JUL-1999; 99US-0361619.  
XX  
PA (CONN-) CONNAUGHT LAB LTD.  
XX  
PI Loosmore SM, Sasaki K, Yang Y, Klein MH;  
XX  
DR WPI; 2001-159722/16.  
XX  
PT P-PSDB; AAB69133.  
XX  
PS New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
XX useful in protective vaccines and for diagnosis  
XX  
PS Example 3; Fig 2A-W; 247pp; English.  
XX

CC The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. CC The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II) and its truncated versions, are used as immunogenic compositions and vaccines to protect against M. catarrhalis infections, particularly otitis media in humans. (II) is also used as antigen in immunoassays for detecting specific antibodies (Ab), and to generate Ab. (I) are used for recombinant production of (II) and its fragments are used as probes for identifying/cloning 200 kDa protein genes from other strains, and for diagnostic detection of M. catarrhalis. (I) makes possible production of large amount of recombinant immunogens. CC Expression of truncated versions of (II) reduces toxicity of the protein towards the Escherichia coli host. The present sequence represents the M. catarrhalis strain 4223 lambdaDBMBL3 clone 200Kba gene, which is used in the exemplification of the present invention.

XX  
SQ Sequence 6973 BP; 2265 A; 1555 C; 1533 G; 1620 T; 0 other;

Query Match 37.2%; Score 2580.6; DB 22; Length 6973;  
Best Local Similarity 66.9%; Pred. No. 0;  
Matches 4803; Conservative 0; Mismatches 1104; Indels 1270; Gaps 29;

QY 1 atgaatcacatctataaagtcattctttaaaagccacagggcacacattttatggccgtggca 60  
DB 542 atgaatcacatctataaagtcattctttaaaagccacagggcacacattttatggcggtgca 601  
QY 61 gaggcgccaaatccacagc-ggaggagtagcagtagtacgcagcagcaggtggcgag 119  
DB 602 gaggcgccaaatccacagcgggggggggttagctgtgtacagcgcaagtcggcgag 561  
QY 120 ctctctgtctatcgctgactgctgtgcccagcgtcgctatctctgtgtgctggtgctgac 179  
DB 662 tgtatgcactctgagctttgccggtatgcccgcgtcgctgctctgctggtgctgac 721  
QY 180 gctcaatggcagtgcttctgctcaaa-----ataatagcaagatcgctattggt----- 228  
DB 722 gctcagtggcagtgcttctgctcaaaagagatatacaaacatcgcatcgatgggtgaaaca 781  
QY 229 ----accacagcacaatgacatgctcgcttagcaatgaagcattcctgctattggt 284  
DB 782 aaacagcgaagcagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 344  
QY 285 tagtcttgctaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 344  
DB 842 tgaataatgctaacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 898  
QY 345 tgcataatcagcgggctaaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 404  
DB 899 tgcataatcagcgggctaaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 958  
QY 405 catcggtggtgagtgactggtgaggtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 464  
DB 959 catcggtggtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtg 1018  
QY 465 atatttgataggaatagcactaactcaatataatccaaatgg----tctcttagcactct 521  
DB 1019 acatttgcttgatcagcagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1078  
QY 522 tattcaaaaccatagctattacgcgaatcacgagactcaaatggtttctca---gaaata 578  
DB 1079 tattcaacggccatgagctattcaaaagaaataacgagctcaaaaggaataatgagtgaaata 1138  
QY 579 tagacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 638  
DB 1139 tagacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1198  
QY 639 gggctattttgcaacgcctttgtgtacacggttcaacagcgtgaagcagcagcagcagcagcagc 698  
DB 1199 gggctatttttcaacgcctttgtgtacacggttcaacagcagcagcagcagcagcagcagcagc 1258  
QY 699 agtaggtcttacgcgcgaagcgcgaagaaaggaatacaacatcgctattggttcaatgcaaca 758  
DB 758

DB 1259 agtgggtcttgcgcgcacagccgagggccaatctacaatcgctattggttctgtatgcgaac 1318  
QY 759 agctatcaattatggagcactagcccttggtgcagatactcagctgagttgattgggatacgg 818  
DB 1319 atctagctcgttgggagcgatagcccttgggtcaggtactcgtcgtcagctacagggcgag 1378  
QY 819 tattgcctaggttattggttctcagatcccttaataataataataataataataataataataata 875  
DB 1379 tattgcctaggttcaagggtctgtgtctcactcagagtgataataatctctagacggcccta 1438  
QY 876 agcctatgtaccagaaggttaagtgggtcaaacatacaaatcgtctaaagccacgcgcgaatgg 935  
DB 1439 tacaccaaataccacagcagcactagacccccaagttccaagccaccaataataacgaagcggg 1498  
QY 936 ttattttccattggttagtagcactatcaagcgtaaaatcactaatgtcgggtgcaggtta 995  
DB 1499 tccactttccattggttagtaactctatacaaacgtatacaaatcactaatgtcgggtgcaggtgt 1558  
QY 996 tgaggataccgtagcgggtcaatgtggcacagcagctaaagcgggtggagaaatctggcta---a 1052  
DB 1559 taataaaacccgagtcgggtcaatgtggcacagcagctagaagcgggtggaagtgggctaaagga 1618  
QY 1053 gcgtcaaatctacttttaagggtgtagataacgggtactggtgttaagaaaaaacctgggcca 1112  
DB 1619 gcgtagaatctactttcaggtgtagataacggtactgacgtataaaataggtttggataa 1678  
QY 1113 gactttaaccattaaagggtggtgagaccccaagcggacacagcagcagcagcagcagcagcagc 1172  
DB 1679 tactttaactattaaagggtggtg---cagagacccaacgcttaccgagataataataatcgg 1735  
QY 1173 tgggtggttaacagataataactggtctgaaagttaaaacttgcataaaacctaagcgg 1232  
DB 1736 tgtgtgtaaaagagcgtgataaataagtgctcigaaagttaaaacttgcataaaactttaacaa 1795  
QY 1233 tcttgaaacaglttagccacaaaaacccctaaaccgcagcggagaaagtacgtag---gtag 1289  
DB 1796 tcttactgagtgagtaacactacactataatggccacacacagcagtgtaaggtaggtagcag 1855  
QY 1290 tggtaataaacccctgagctacaaacgggtgggtttaaaccctttacc---caacaac 1343  
DB 1856 tagtgactactcagcgtgaattatgagtgatgtagttacttaacccttaccagcccaatacagg 1915  
QY 1344 aaatgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1403  
DB 1916 cagtcag 1975  
QY 1404 taatcgcgacttgaa---gatactactcgtatcaccaaaagataaaaattggttttagcaa 1460  
DB 1976 agaaacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2035  
QY 1461 taaagctggtagcagttgatgaaac 1516  
DB 2036 agatggtgagttgatgaaac 2095  
QY 1517 ----- 1517  
DB 2096 tagtgttgcattaccatagacaaatggcaggtatgtagcaggttaataaaagatcagtaact 2155  
QY 1518 tggcaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1577  
DB 2156 tggcaaaagtagcagtgctaaacagtcggttaccatcgacagcagcagcagcagcagcagcagc 2215  
QY 1578 taacaaaatccaaagtgggtgctgagtgagcagcagcagcagcagcagcagcagcagcagcagcag 1637  
DB 2216 tactttaacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2275  
QY 1638 aatgcgcgcaaaattcggcactactcgttattaccgagaggaatattggttttgcgtgagc 1697  
DB 2276 gagtggcaatgtacgcgcac 2335  
QY 1698 tgatggtaaaagtgtgataa----- 1715  
DB 2336 tgaaggcactagtgtgataaaatttagtgttaagggttagtggtacgaaacaatagcttagttac 2395





||||| 4508 aaccattgaagttaa---gataaaaaacttgcgtaaaaaaccacacacattgaccagtac 4564  
QY 3849 aagtctaattgaatgca---acaaaatttagtgcgccgaatggcgatgccttggttaa 3905  
Db 4565 tggcacaggtgctaataatttggcctaagcaatacaagctactgctgcatgctgtgtca 4624  
QY 3906 agccagtgtatcgcacacccatctaaatccttgctgctgacatcccaaacgcgcaagg 3965  
Db 4625 gscagtgatcgttgctcaactcaaaccttactcggcagacatccaaactgccaagg 4684  
QY 3966 ggcagccaagcaagcagctcagcaagctatgtgtagctgtgatggtgcaacaggtcatcta 4025  
Db 4685 ggcagccaagcgaacactcagcgaggtatgtgtagctgtgatggtgcaataaggtcatcta 4744  
QY 4026 tgacagtcaccgataagaagctactatcaagtcagtacaaggtcagtgggacaaacaa 4085  
Db 4745 tgacagtcaccgataagaagctactatcaagtcagaaggtcagtgggacaaacaa 4804  
QY 4086 agaagttgccaaagacaactggtgcgccaagcccaaaccccgatggcacttggtctca 4145  
Db 4805 agaagttgccaaagacaactggtgcgccaagcccaaaccccgatggcacttggtctca 4864  
QY 4146 aatgaatgcataatcagtcatttaacaaagagcaagtaaatgtagtgcataaaaaagcaagg 4205  
Db 4865 aatgaatgcataatcagtcatttaacaaagagcaagtaaatgtagtgcataaaaaagcaagg 4924  
QY 4206 catcaatgaagacaacgcctttatcaaaaggcttgaaacgcgcgaagacaccaaacc 4265  
Db 4925 catcaatgaagacaacgcctttgtaaaagacttgaaaaagcgcctttgtataacaaac 4984  
QY 4266 caaaacgcgcgtaactgtgtgtgatttaaatgccgttgccaaacacccgtgacctt 4325  
Db 4985 caaaacgcgcgtaactgtgtgtgatttaaatgccgttgccaaacacccgtgacctt 5044  
QY 4326 tgacgggagatacagcacacgcgttaaaaaactggcgagactttgacatcaagggtgg 4385  
Db 5045 tgacgggagatacagcacacgcgttaaaaaactggcgagactttgacatcaagggtgg 5104  
QY 4386 gcaaacagacacccaataaagctaacccgataataacatcgggtggtgagcgggtactgtatgg 4445  
Db 5105 gcaaacagacacccaataaagctaacccgataataacatcgggtggtgagcgggtactgtatgg 5164  
QY 4446 ctctcactgtcaaaacttgcaagagcctaacaactttaacagcgtttaatgcagggtggac 4505  
Db 5165 ctctcactgtcaaaacttgcaagagcctaacaactttaacagcgtttaatgcagggtggac 5224  
QY 4506 cagaattgatgaaaaagcgtctctttttagacgcgaacgcgtcgaagccaaagcaaacac 4565  
Db 5225 caaaattgatgaaaaagcgtctctttttagacgcgaacgcgtcgaagccaaagcaaacac 5284  
QY 4566 cctgtgtctaagtgccaaatgggctggacctgggtggcgaacgcgtatcgaatcgggtgc 4625  
Db 5285 cctgtgtctaagtgccaaatgggctggacctgggtggcgaacgcgtatcgaatcgggtgc 5321  
QY 4626 agctgttgatgataacgtagcgggtgaactttaagcaggtttaatgaagttgccaaacggt 4685  
Db 5322 ----- 5321  
QY 4686 caacaacccataaacacaaagtaactcagggtgcgtcattaccctttgtgtgtaaccgatgc 4745  
Db 5322 ----- 5321  
QY 4746 caatggcaagcccatacaatggcaccgatggcaagcccccaaaaagccatcaagggtgcga 4805  
Db 5322 ----- 5321  
QY 4806 tggtaatactactacgcgaacccacacgcgcgtacctgtggcaaaagatggcaagcccat 4865  
Db 5322 ----- 5321  
QY 4866 caccgatgcggacaaaacttgccaatctggcagctcatgtgcaaaccccttgatgcagggtca 4925

Db 5322 ----- 5321  
QY 4926 tcaagtgtggcaagccttagggcgaactcagatgccaatcccttaaccacacatcaagtc 4985  
Db 5322 ----- 5321  
QY 4986 cactttgccacaaattgacacacacacacaggttaatgccaatgcagggtgcaagcccaag 5045  
Db 5322 ----- 5321  
QY 5046 tctgccagcctatcagcagcacagcaaaagtaatgctgccagtgctcaaaagatgtctaaa 5105  
Db 5322 ----- 5321  
QY 5106 ttaggtcttaacttgcagaccaatcaaatcaagtgactttgtcnaaagcctatgatcac 5165  
Db 5322 ----- 5321  
QY 5166 cgtcaactttgtcaatggtacaggtgcgacatcacaaagcgtgcgttagctgatggcac 5225  
Db 5322 ----- 5321  
QY 5226 gatgagtaacatcacctgcacacccgcttagcagcagcagatgatgagcgaatgtgct 5285  
Db 5322 ----- 5321  
QY 5286 tatcaaaagcgaagatgtagtaattctacaaagcagacacactcatgcnaaacgcgtcact 5345  
Db 5322 ----- 5321  
QY 5346 aaaagcaggaataatcagcagtgatgcgaactccaaactcctaagcctgtctaaagccttgaaccc 5405  
Db 5322 ----- 5321  
QY 5406 caatgctgttaaagcagctacagcgatgcagtgcttctaataacttatcaaaagcgt 5465  
Db 5322 ----- 5321  
QY 5466 atttaaatccaaagatggtacaactactaccacagtaagctctgtagcagtatcca 5525  
Db 5322 ----- 5321  
QY 5526 aggcgaagataaacagcagcgtaccccttaagcaagaatgggctgaatgtaggcgttaaggt 5585  
Db 5322 ----- aaggt 5326  
QY 5586 catcagcaatgtgggttaaaggcacaacaaagacaccacgcgtgccaatgtacaacagttaa 5645  
Db 5327 catcagtaatgtggcgaaggcacaacaaagataccgcgcgtgccaatgtacaacagttaa 5386  
QY 5646 cgaagtacgcaactgttgggtcttggtaattgctgtaattgataacgcctgacgcgcaatca 5705  
Db 5387 cgaagtacgcaactgttgggtcttggtaattgctgtaattgataacgcctgacgcgcaatca 5446  
QY 5706 ggtacaactgcccgcacatacaaaaagaccccaatcaggttcatcatcttaacgcgcactgt 5765  
Db 5447 ggtacaactgcccgcacatacaaaaagaccccaatcaggttcatcatcttaacgcgcactgt 5506  
QY 5766 catcaaaagcagcagcgttacttggcggtaaaagtaataacgataccgaaaaacttgcac 5825  
Db 5507 catcaaaagcagcagcgttacttggcggtaaaagtaataacgataccgaaaaacttgcac 5566  
QY 5826 tgggtgtgtacaagtggcgctggtataaagacgcgcacgcgtcaacgcgttaacgcaatgt 5885  
Db 5567 tgggtgtgtacaagtggcgctggtataaagacgcgcacgcgtcaacgcgttaacgcaatgt 5626  
QY 5886 ttgggtcaaaaccccaaaagatggcagcaaaaaagccctgtctgcgaactttaaacgcgc 5945  
Db 5627 ttgggtcaaaaccccaaaagatggcagcaaaaaagccctgtctgcgaactttaaacgcgc 5686  
QY 5946 aggtcagacaaactatttgacacaaacccgcgaagcattgacagataaataaagaa 6005  
Db 5687 aggtcagacaaactatttgacacaaacccgcgaagcattgacagataaataaagaa 5746







Db	4615	-----	4614	Db	4726	gctgacgcaatcaggtaaacatgcccagacatcaaaaaagaccccaaatcagggttcac	4785
Qy	4672	gttgccaaacggtcaacaacctaatacaacccaagtaactcaggtgcttattacccttt	4731	Qy	5752	tctaacccgactgtcatcaaaagcaggcacggtacttgccgttaaaagtaataacgatacc	5811
Db	4615	-----	4614	Db	4786	tctaacccgactgtcatcaaaagcaggcacggtacttgccgttaaaagtaataacgatacc	4845
Qy	4732	gttgtaacgagtccaatggcaagcccatcaatgacaccggtgcaagcccaaaaagcc	4791	Qy	5812	gaataactggcactggttggtgtacaagtgggctgggttaaaagcagcgaacgcgtacagcc	5871
Db	4615	-----	4614	Db	4846	gaataactggcactggttggtgtacaagtgggctgggttaaaagcagcgaacgcgtacagcc	4905
Qy	4792	atcaagggcgccgagtgttaatactatcacgccaacgccaacggtgtacctgtggacaaa	4851	Qy	5872	gatttaagcaatgtttgggtcaaaaccccaaaaagatggcagcaaaaaagccctgctgcc	5931
Db	4615	-----	4614	Db	4906	gatttaagcaatgtttgggtcaaaaccccaaaaagatggcagcaaaaaagccctgctgcc	4965
Qy	4852	gatggcaagcccatccaccgatcggaacaaattgcgaacttggcagctcatggcaaaccc	4911	Qy	5932	acttataacgcgcaggtcagaccactatttgaccaacaaccccgagagaagccattgac	5991
Db	4615	-----	4614	Db	4966	acttataacgcgcaggtcagaccactatttgaccaacaaccccgagagaagccattgac	5025
Qy	4912	cttgatgcaggtcatcaagtggtggcaagccttagggcggaactcagatgccatccctta	4971	Qy	5992	agaataaatgaacaaaggtatccgcttctccatgttcaacagatggcaatcaagagcctgtg	6051
Db	4615	-----	4614	Db	5026	agaataaatgaacaaaggtatccgcttctccatgttcaacagatggcaatcaagagcctgtg	5085
Qy	4972	accaacatcaagtcacacttggccaaaattgacacacaaacacaggttaatgccaatgca	5031	Qy	6052	gtacaagggcgtaacgcgcatctgactcaagtgcctcaggcaacgcaactcagtgcgataggt	6111
Db	4615	-----	4614	Db	5086	gtacaagggcgtaacgcgcatctgactcaagtgcctcaggcaacgcaactcagtgcgataggt	5145
Qy	5092	aaagatgtctaagttaggctttaacttgcagaccatcacatcaatcaagtggactttgtc	5151	Qy	6112	ttccaggcccaagcgagatggtgaagccgcgttgcacataggcagacacaaacccagcagcc	6171
Db	4615	-----	4614	Db	5146	ttccaggcccaagcgagatggtgaagccgcgttgcacataggcagacacaaacccagcagcc	5205
Qy	5152	aaagcctatgataccgtcaacttctcaatggttacaggtgcgacatcacaaagcgtgct	5211	Qy	6172	aacaaatccatccgcatcggtgataacgcgaacagccacgcggcgatcaatccatcgccatc	6231
Db	4615	-----	4614	Db	5206	aacaaatccatccgcatcggtgataacgcgaacagccacgcggcgatcaatccatcgccatc	5265
Qy	5212	agtgctgtaggcacgatgagttaacatcacccgtcaacaccgcttagcagcgccgatgat	5271	Qy	6232	ggtacagggcaatgtgtgtaacaggtaagcactctgtgtgccaactcggcgacccacgactgtt	6291
Db	4615	-----	4614	Db	5266	ggtacagggcaatgtgtgtaacaggtaagcactctgtgtgccaactcggcgacccacgactgtt	5325
Qy	5272	gatggcaatgtgcttatcaaaagccaaagatggttaagttctacaaagcagacacctoatg	5331	Qy	6292	aaggtgtataacagttacagtggtggttaataacaaacagtttctcgtgccaactcagacc	6351
Db	4615	-----	4614	Db	5326	aaggtgtataacagttacagtggtggttaataacaaacagtttctcgtgccaactcagacc	5385
Qy	5332	ccaaacggctcaactaaaagcaggcaaatcagccagtgatgccaaaactccaactggctta	5391	Qy	6352	gatgtcttgggtgggcaataacataccgtgacccggaagtaactcgttgccttaggt	6411
Db	4615	-----	4614	Db	5386	gatgtcttgggtgggcaataacataccgtgacccggaagtaactcgttgccttaggt	5445
Qy	5392	agccttgtaaaccctaatgctggttaaaaggcagttacaggcgatgcagttgctttaaac	5451	Qy	6412	tcaaaactctgcacatcagtcagggcacacacgagcacaacagcccaaaaaatctgaagcc	6471
Db	4615	-----	4614	Db	5446	tcaaaactctgcacatcagtcagggcacacacgagcacaacagcccaaaaaatctgaagcc	5505
Qy	5452	ttatcaaaagcggtatttaattccaaagatggtacaactactacacagtaagctctgat	5511	Qy	6472	acagcagggtacaaacccacacagcaggtgcaacaggtacgggttaaaagccttgcctggacaa	6531
Db	4615	-----	4614	Db	5506	acagcagggtacaaacccacacagcaggtgcaacaggtacgggttaaaagccttgcctggacaa	5565
Qy	5512	ggcatcagttccaagggcaaaagataaacagcagcatcaccccttaagcaaaagatgggtgaat	5571	Qy	6532	acggcggttgggtgcggtctccgtggtggtcctcaggtgctgaacgcctgataccaaaatgtg	6591
Db	4615	-----	4614	Db	5566	acggcggttgggtgcggtctccgtggtggtcctcaggtgctgaacgcctgataccaaaatgtg	5625
Qy	5572	gtaggcggttaaggtcatcagcaatgtgggttaaaaggcacaacaaagacacagcgtccaat	5631	Qy	6592	gcagcaggtgaggtcagtgccaccacacagcaggtgcaatgggttaaaagccttgcctggacaa	6651
Db	4615	-----	4614	Db	5626	gcagcaggtgaggtcagtgccaccacacagcaggtgcaatgggttaaaagccttgcctggacaa	5685
Qy	5632	gtacaacagttaaacaaagtacgcaacttgggttcttgggttaaatgctggttaataataac	5691	Qy	6652	gcaccccaaaagcattgccaaacgcaacaaatggagcttgacatctgataccacaaacgaa	6711
Db	4615	-----	4614	Db	5686	gcaccccaaaagcattgccaaacgcaacaaatggagcttgacatctgataccacaaacgaa	5745
Qy	5692	gctgacggcaatcagggttaacattgcccagatcacaaagaccccaaatcagggttcac	5751	Qy	6712	aataaaagcaatgagggatttcatcagcaggtggtggtggtggtggtggtggtggtggtggtg	6771
Db	4615	-----	4614	Db	5746	aataaaagcaatgagggatttcatcagcaggtggtggtggtggtggtggtggtggtggtggtg	5805
Qy	5752	gctgacggcaatcagggttaacattgcccagatcacaaagaccccaaatcagggttcac	5805	Qy	6772	attcctggcagatccatggtttaccgggggtatggcaccacacacggtcaggtggtggtggtg	6831
Db	4615	-----	4614	Db	5806	attcctggcagatccatggtttaccgggggtatggcaccacacacggtcaggtggtggtggtg	5865

QY 6832 gcagtgaggactgtcgaagctgtcggataatgtgccaatgttgtaattattataaatacaatggttca 6891  
|||||  
Db 5866 gcagtgaggactgtcgaagctgtcggataatgtgccaatgttgtaattattataaatacaatggttca 5925  
QY 6892 gccgatacccaagcccatgtaggcgccgagtggtggtcaggtgtttcaacttt 6942  
|||||  
Db 5926 gccgatacccaagcccatgtaggcgccgagtggtggtcaggtgtttcaacttt 5976

## RESULT 9

AAF59106  
ID AAF59106 standard; DNA; 6259 BP.  
XX  
AC AAF59106;  
XX  
XX 24-APR-2001 (first entry)  
XX  
XX M. catarrhalis M56 200kDa gene in pKS348 SEQ ID NO:12.  
DE  
XX  
XX Moraxella catarrhalis strain Q8; major outer membrane protein;  
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection; ds.  
XX  
OS Moraxella catarrhalis.  
XX  
XX WO200107619-A1.  
XX  
XX 01-FEB-2001.  
XX  
XX 26-JUL-2000; 2000WO-CA00870.  
PF  
XX  
XX 27-JUL-1999; 99US-0361619.  
PR  
XX  
XX (CONN-) CONNAUGHT LAB LTD.  
PA  
XX  
PI Loomore SM, Sasaki K, Yang Y, Klein MH;  
XX  
XX WPI; 2001-159722/16.  
DR  
DR P-PSDB; AAB69127.  
XX  
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
PT useful in protective vaccines and for diagnosis  
PT  
XX  
XX Claim 1; Fig 8A-V; 247pp; English.  
XX  
XX The present invention describes an isolated and purified nucleic acid (I)  
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.  
CC The 200 kDa outer membrane protein (II) has antibacterial activity and  
CC can be used in vaccines. (II), and its truncated versions, are used as  
CC immunogenic compositions and vaccines to protect against M. catarrhalis  
CC infections, particularly otitis media in humans. (II) is also used as  
CC antigen in immunoassays for detecting specific antibodies (Ab), and to  
CC generate Ab. (I) are used for recombinant production of (II) and its  
CC fragments are used as probes for identifying/cloning 200 kDa protein  
CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
CC (I) makes possible production of large amount of recombinant immunogens.  
CC Expression of truncated versions of (II) reduces toxicity of the protein  
CC towards the Escherichia coli host. The present sequence represents the  
CC M. catarrhalis M56 200kDa gene in pKS348, which is given in the  
CC exemplification of the present invention.  
XX  
XX Sequence 6259 BP; 2067 A; 1414 C; 1393 G; 1385 T; 0 other;

Query Match 35.8%; Score 2484.8; DB 22; Length 6259;  
Best Local Similarity 66.7%; Pred. No. 0;  
Matches 4674; Conservative 0; Mismatches 1067; Indels 1269; Gaps 28;

QY 167 tgatcgtgcagcgtcaatgtgagtgcttatgtctcaataatagca-----agatcg 220  
|||||  
Db 2 tgatcgtgcagcgtcagtgagtgcttatgtctcaataatagca-----agatcg 61

QY 221 catttgtt-----accacaggccaacatgacatgctcctcggctagcaataagcaat 271  
|||||  
Db 62 caattgtgaacaaaaccagccaagcgtcagcactgccaagcgagcgtgtatcgag 121  
QY 272 ccattgtattgtctgtctgaagccacatgccaatcaagctatgtctatcgttgta 331  
|||||  
Db 122 ccattgtctatgttgtaaaatgctaaccacagggcggtcaagccatcgcctcgttgta 181  
QY 332 gcaaacaccagatcctcgttaataacagcggctaatcagaaggcaggttcccacgccaaaagta 391  
|||||  
Db 182 gtaa--taaaactgtcaatggaagcagtttggtataagataggtaccgatgtcagggtc 238  
QY 392 aagaagtcacatgcgcacatcgggtgtgattactggtcgggtgtgctcgtatgccattg 451  
|||||  
Db 239 aagaagtcacatgcgcacatcgggtgtgattactggtcgggtgtgctcgtatgccattg 298  
QY 452 gtatgtatgacttatatttggtataggaatgacactaacttaataatcccaatgg---tc 508  
|||||  
Db 299 gtatgtatgacttatatttggtataggaatgacactaacttaataatcccaatgg---tc 358  
QY 509 ttcttgacactcttattccaaaaccatacagctatttacgccaatacagactccaatggt 568  
|||||  
Db 359 tgattaacgactcttattacacggccatcagctatttaaaagaatacgaagctcaagata 418  
QY 569 ctca---gaaatatagacgcacagcagcagaagacacgcgcagctactgagtgagcca 625  
|||||  
Db 419 atgatgtaaaaatagacgcacacaccgcaagcgcagcgcagctactgagtgagcca 478  
QY 626 tggcatatgcaaaaggtcattttgccaacgctttgttacacggtcaacagctcaagcca 685  
|||||  
Db 479 tgcataatgacacagggctcattttccacgctttgttacacggtcaacagctcaagcca 538  
QY 686 actattccttggtcagtaggtcttaccgccaagcgaagaggaatacacaatcgtctattg 745  
|||||  
Db 539 cctattccttggtcagtaggtcttaccgccaagcgaagaggaatacacaatcgtctattg 598  
QY 746 gttcattgcaacagctatcaattatggacactagcccttggtgcagatcactcgagttg 805  
|||||  
Db 599 gttcattgcaacagctatcagctgttggagcgtatgacctggtgcaggtactgctgctc 658  
QY 806 atttggtattcaggtatgctcctaggttattgtctcagatccttaataataataata 865  
|||||  
Db 659 agctacagggcagctatgctcctaggttcaaggtctgttgcactcagatgataataat 718  
QY 866 ata---ataataagcctatgtaccagaaggtaatggtggtcaaacataataatcgtctaaag 922  
|||||  
Db 719 ctgacacggcgtatcacacaaataccacaggcactagaccacagtttcaagccaata 778  
QY 923 ccacggcgaatggtttatttccattggtatgacactatcaagcgtataaatcatcaatg 982  
Db 779 atacgaagggcggttccactttccattggtatgactctatcaaacgtataaatcatcaatg 838  
QY 983 tcggtgcaggttatgaggtacacgctcgttcaatgtggtcagcagctaaaagcgtgagga 1042  
|||||  
Db 839 tcggtgcaggttataaaacacagctcgttcaatgtggtcagcagctaaaagcgtgagga 898  
QY 1043 atctgggcta---agcgtcaaatcttttaagggtgatgataacggtactgctgcttaaga 1099  
|||||  
Db 899 agtgggctaaggagcgtgagaaattactttcaggtgatgataacagctactgactgataaaa 958  
QY 1100 aaaaactggcgagacttttaaccattaaaagtggtgagaccccaagcggaacgtaacccg 1159  
|||||  
Db 959 taggtttggataataactttaactataaagtggtg---cagagaccaacgcattaaccc 1015  
QY 1160 ataataaactgtgtgtggttaacagataataactgtgtcctgaaggttaaaccttcta 1219  
|||||  
Db 1016 ataataatcgtgtgtgtaaaagagcgtgataatagtggtcgtgaaggttaaaccttcta 1075  
QY 1220 aaaacacctgaagcgtgtgaaacaggttagcacccaaaaaacactaacccgcgcgagaaagta 1279  
|||||  
Db 1076 aaactttaaacactcttactgaggtgataacactataatgataacacacacacagta 1135  
QY 1280 cggtag---gtagtggttaataaacaccgctgagctacaaaagcgttggtttaacctttacc 1336





Qy	3302	ttagtaaccttactccagttatgacacactcaaaagacctctgatgtcatcaacctttgcag	3361
Db	3308	ttagtaaccttactccagttatgacacactcaaaagacctctgatgtcatcaacctttgcag	3367
Qy	3362	gtgaaacggcattaccaccaaggtaaataaaaggtgtgtgtcgtgtgggcattgaccaaa	3421
Db	3368	gtgaaacggcattaccaccaaggtaaataaaaggtgtgtgtcgtgtgggcattgaccaaa	3427
Qy	3422	ccaaaggcttaaccagccttaagctgaacgtgtgggttaataataatgtggcaaggcaattgtca	3481
Db	3428	ccaaaggcttaaccagccttaagctgaacgtgtgggttaataataatgtggcaaggcaattgtca	3487
Qy	3482	ttgacagtaaagatgttcaaaataccatcacaggactaagcaacacitctagcttaatttta	3541
Db	3488	ttgacagcctaaatgttcaaaataccatcacaggactaagcaacacitctagcttaatttta	3547
Qy	3542	ccaatga-----tgggtgaggacacgcactaagccaaggcttgcctatgacaccgca	3595
Db	3548	ccaatgataaaggtagcgtacgcaccacagaacaggccaataataatcaagacgaagaca	3607
Qy	3596	aaaccgctgcgcgcagcatgtgtgtatgtgtctaaacgcaggctttaacttgcgaaggcaatg	3655
Db	3608	aaaccgctgcgcgcagcatgtgtgtgtctaaacgcaggctttaacttgcgaaggcaatg	3667
Qy	3656	gtgaagcggttgactttgtctccacttatgacacttgttgactttatcgatggcaatgccca	3715
Db	3668	gtgaagcggttgactttgtctccacttatgacacggtcaacttggcagatggcaatgccca	3727
Qy	3716	ccacgcctaaggtagcctatgatgacacaagcaaaacacagtaaaagtggtctatlgatgtca	3775
Db	3728	ccacgcctaaggtagcctatgatgacacaagcaaaacacagtaaaagtggtctatlgatgtca	3787
Qy	3776	atgtgtaataaaaaaccattgaagtgacaagtgtataaaaaacttggcgtcaaaaccacca	3835
Db	3788	atgtgtagatatacaacattgaagttaa- --gataaaaacttggcgtataaaaccacca	3844
Qy	3836	cactgacaaaacaagtgctaatgttaatgca- --accaaatttagtgcgcgcgatggcg	3892
Db	3845	cattgaccagctactggcaggtgtctataataatttgccttaagcaatacagctactgtggcg	3904
Qy	3893	atgccttgtttaagccagtgatatgcgcacccactataaacttggcttggcgacatcc	3952
Db	3905	atgccttgtcgaagccagtgatatcgttgcctatcaaaacaccttatcttggcgacatcc	3964
Qy	3953	aaaccgcaaaaggggcaagcccaagcaagcagctcagcaagctatgtggatgctgatggca	4012
Db	3965	aaactgcaaaaggggcaagcccaagcaacactcagcaggtatgtggaatgctgatggca	4024
Qy	4013	acaaggtcatctatgacgtaccgataagaagtactatcaagtcaaatgacaagggtcaag	4072
Db	4025	ataagggtcaactatgacgtaccgataagaagtactatcaagtcaaatgacgaagggtcaag	4084
Qy	4073	tggcaaaaacaaagaagtgtccaaagcaaaactggttcgcgccaaagcccaaaccccgatg	4132
Db	4085	ttgataaaacaaagaagtgtccaaagcaaaactggttcgcgccaaagcccaaaccccgatg	4144
Qy	4133	gcacattggtctaaatgtaatgtcaaatcagttcaatacaagacgaagtaaatgatgtcca	4192
Db	4145	gcacattggtctaaatgtaatgtcaaatcagttcaatacaagacgaagtaaatgatgtcca	4204
Qy	4193	ataaaagcaagcatcaatgaagcaaacgcctttatcaaggcttgaaaagccgcgcca	4252
Db	4205	ataaaagcaaggtcaatgaagcaaacgcctttttaaaggacttgaaaaagccgctt	4264
Qy	4253	aagacacaaaacccaaaacgcgcgagtaactgtgtgggtgatttaaatgccgttgcacaaa	4312
Db	4265	ctgatacaaaaacccaaaacgcgcgagtaactgtgtgggtgatttaaatgccgttgcacaaa	4324
Qy	4313	caccgctgacctttgacggggatagacaggcaaacgcgtataaaaacttggggagacatttga	4372
Db	4325	caccgctgacctttgacggggatagacaggcaaacgcgtataaaaacttggggagacatttga	4384

QY	4373	ccatcaaaagtgggcaaacacagacacacaataaagctaaccgataataacatcgggtgtggtag	4433
Db	4385	ccatcaaaagtgggcaaacacagacacacaataaagctaaccgataataacatcgggtgtggtag	4444
QY	4433	cagggtactgatggcttcaactgctcaaaacttgccaaagacctaaccatctttaaagcggtta	4492
Db	4445	cagggtactgatggcttcaactgctcaaaacttgccaaagacctaaccatctttaaagcggtta	4504
QY	4493	atgcagggtggcaccagaattgatgaaaaggcatctctttttgtagacgcgaacgggtcaag	4552
Db	4505	atgcagggtggcaccagaattgatgaaaaggcgctgctctttttgtagactcaagcggtcaag	4564
QY	4553	ccaaagcaaacacccctgctgctaagtgccaatgggctggacctggggtggcacaacgcatac	4612
Db	4565	ccaaagcaaacacccctgctgctaagtgccaatgggctggacctggggtgggc-----	4614
QY	4613	gtaacatcgggtgcagctgttgatgataaagatcggtggaactttaagcagtttaataagag	4672
Db	4615	-----	4614
QY	4673	ttgccaaaacggttcaaacacctaataaacacaaagtaactcaggtgcgtcattaccctttg	4732
Db	4615	-----	4614
QY	4733	tggtaaccgatgccaatggcaagcccatcaatctggcaccgatggcaagcccccaaaagcca	4792
Db	4615	-----	4614
QY	4793	tcaaggggccgatggtataatactatcacgcgccaacgcgctacacttgtggacaag	4852
Db	4615	-----	4614
QY	4853	atggcaagccatcacccgatcgggacaaaacttgccaactctggcagctcatggcaaacccc	4912
Db	4615	-----	4614
QY	4913	tigtatgcaggtcatcaagtgggtggcaagccagcggaactcagatgccaatcacccctaa	4972
Db	4615	-----	4614
QY	4973	ccaacatcaagtccactttggccacaaatgacacaccaacacaggttaatgccaatgcag	5032
Db	4615	-----	4614
QY	5033	ggcaagcccaaaagtctgccagcctatcagcagcacagcaagtaattgctgcagtgctca	5092
Db	4615	-----	4614
QY	5093	aagatgtctaaatgtagcctttaacttgcagaccaatcacatcaagtggactttgttca	5152
Db	4615	-----	4614
QY	5153	aagcctatgataccgtcaactttgtcaatggtacaggtgccgacatacaagcggtgcgta	5212
Db	4615	-----	4614
QY	5213	gtgctgatggcagatgagtaacatcacogtcaacaccgcttagcagcgacgatgatg	5272
Db	4615	-----	4614
QY	5273	atggcaatgtcttatcaagcccaagatggtgaatttctacaagcagacgacctcatgc	5332
Db	4615	-----	4614
QY	5333	caaacggtcactaaaagcaggcaaatcagccagtgatgccaaaactccaactgggtctaa	5392
Db	4615	-----	4614
QY	5393	gccttgttaacccaatgctggtaaaggcagttacagcgcatgagtggtctttaataact	5452
Db	4615	-----	4614
QY	5453	tatcaaaagcgggtattttaaatccaaagatggtatacaactaccacagtaagctctaatg	5512





Thu Sep 13 14:17:52 2001

```

Db 2924 tttgctgcaacgagcgacggtgctgtctgctgagcgcaagcgcgaggaagacgt 2983
QY 6580 atccaaatgtgagcgagcggtgagtgagtgccaccagcaccgagcggttcaatggtagc 6639
Db 2984 atccaaacgctcgcgagcgaggaatttccgccacttccaccgagtgaggttaacggcagc 3043
QY 6640 cagttgtacaaagccaccgagcggttgcacacgcaacgaatgagcttgaccatcgatc 6699
Db 3044 cagttgtatgtgtgagcaaaagggttaacaa-----atcttgctggacaagt 3091
QY 6700 caccacacgaaataaagcccaatgcaggagtttcatcagcagatggcggtccatg 6759
Db 3092 aataaagtgggcaacgtgcagatgcaggtacagcaagtgattagcagcttcaacagtta 3151
QY 6760 ccaacagctacattctgctgagatccatggttaccggggtattgccaccacacacggt 6819
Db 3152 ccacagcgctctatgcaggttaaatcaatggttctatgcgggaagtagttatcaaggt 3211
QY 6820 caaggtgctgagcgagtgagctgctgaggaatgctgagataatggtcaatgggtattaaa 6879
Db 3212 caaaatggttagctatcggtggtatcacgaatcttcgataatggcaaaagtgattatcg 3271
QY 6880 atcaatggttcagcagataccacagcgcatgtagggcgcgaggtggtgcaggttttccac 6939
Db 3272 ttgtcagggcacaccataagcaggttaaacagcgcttgacagcaggtgtgtggtaccag 3331
QY 6940 t 6940
Db 3332 t 3332

```

RESULT 13

AAA92498  
ID AAA92498 standard; DNA; 3342 BP.

XX AAA92498;

XX 17-JAN-2001 (first entry)

DE Haemophilus influenzae adhesin (Hia) gene from NTHi strain K22.

KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;

KW non-typeable Haemophilus influenzae; antinflammatory; auditory;

KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;

KW diagnosis; immunogenic; antigen; ds.

XX Haemophilus influenzae.

OS Haemophilus influenzae.

XX WO200055191-A2.

XX 21-SEP-2000.

XX 16-MAR-2000; 2000WO-CA00289.

XX 16-MAR-1999; 99US-0268347.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Yang Y, Klein MH;

XX WPI; 2000-618897/59.

XX P-PSDB; AAB23859.

XX Novel nucleic acid encoding Haemophilus influenzae adhesin protein, for  
PT use as antigens and vaccines and for treating Haemophilus influenzae  
PT infection -  
XX Claim 1; Fig 23; 275pp; English.

XX The present sequence represents a Haemophilus influenzae adhesin (Hia)  
CC gene from the non-typeable Haemophilus influenzae (NTHi) strain K22.  
CC Hia genes and proteins have antinflammatory, auditory and antibacterial

QY 6940 t 6940

Db 3023 t 3023

RESULT 12

AAA92495  
ID AAA92495 standard; DNA; 3354 BP.

XX AAA92495;

XX 17-JAN-2001 (first entry)

XX Haemophilus influenzae adhesin (Hia) gene from NTHi strain K9.

DE Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;

KW non-typeable Haemophilus influenzae; antinflammatory; auditory;

KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;

KW diagnosis; immunogenic; antigen; ds.

XX Haemophilus influenzae.

OS Haemophilus influenzae.

XX WO200055191-A2.

XX 21-SEP-2000.

XX 16-MAR-2000; 2000WO-CA00289.

XX 16-MAR-1999; 99US-0268347.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Yang Y, Klein MH;

XX WPI; 2000-618897/59.

XX P-PSDB; AAB23858.

XX Novel nucleic acid encoding Haemophilus influenzae adhesin protein, for  
PT use as antigens and vaccines and for treating Haemophilus influenzae  
PT infection -  
XX Claim 1; Fig 22; 275pp; English.

XX The present sequence represents a Haemophilus influenzae adhesin (Hia)  
CC gene from the non-typeable Haemophilus influenzae (NTHi) strain K9.  
CC Hia genes and proteins have antinflammatory, auditory and antibacterial  
CC activities, and can be used in the production of a vaccine. An  
CC immunogenic composition comprising an Hia gene, a polypeptide encoded  
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
CC protection against disease caused by Haemophilus strains in a  
CC susceptible host, preferably a human. An Hia protein is useful as an  
CC antigen, in immunogenic preparations including vaccines, as a carrier  
CC is useful for treating diseases caused by the infection of Haemophilus  
CC influenzae such as meningitis, epiglottitis, septicaemia and otitis  
CC media. Recombinant production of Hia favours high recovery of the  
CC protein compared to the low recovery of native protein from Haemophilus  
CC influenzae species. A truncated protein has a significantly higher  
CC amount of recovery than a full-length protein.

XX Sequence 3354 BP; 1082 A; 639 C; 866 G; 767 T; 0 other;

Query Match 2.0%; Score 141.4; DB 21; Length 3354;  
Best Local Similarity 57.8%; Pred. No. 6.7e-27;  
Matches 278; Conservative 0; Mismatches 191; Indels 12; Gaps 1;

QY 6460 aaatctgacggcagcaggtacacacacagcaggtgcaacaggtacggttaaaagc 6519

Db 2864 aaatctgagtgacggcggtacacatacaactgctggcacaacggtgacggttaaacgc 2923

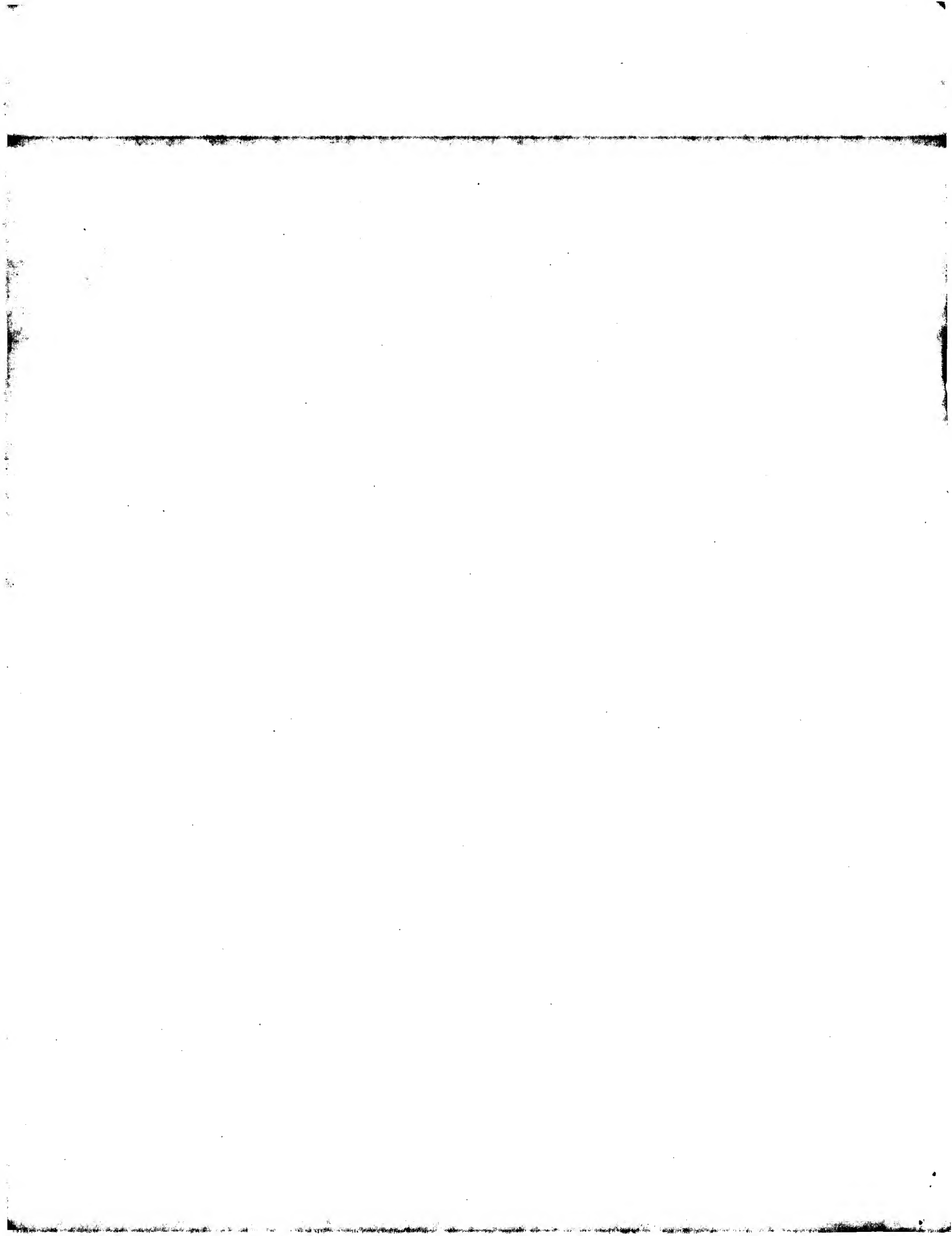
QY 6520 tttgctggacaacgcggtgtgtgctcgtggtgctcgtggtgctcgtggtgctgacgcgt 6579





---





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 12:57:17 ; Search time 454.28 Seconds  
(without alignments)  
2892.925 Million cell updates/sec

Title: US-09-361-619-10  
Perfect score: 6942  
Sequence: 1 atgaatcacatctataaagt.....ttggtgcaggttttcacttt 6942

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2580.6	37.2	6973	1	US-08-478-370-1
2	2382.4	34.3	9542	4	US-08-968-685A-9
3	83.6	1.2	2037	4	US-08-913-942-14
4	82.8	1.2	3294	1	US-08-409-995-1
5	82.8	1.2	3294	3	US-08-685-467-1
6	82.8	1.2	3294	4	US-08-913-942-1
7	76.4	1.1	7291	4	US-08-913-942-3
8	69.8	1.0	5738	1	US-08-409-995-3
9	69.8	1.0	5738	3	US-08-685-467-3
10	58.4	0.8	60	1	US-08-478-370-4
11	56.8	0.8	1797	4	US-09-377-155-12
12	56.8	0.8	1797	4	US-09-377-155-4
13	55.8	0.8	7218	1	US-08-232-463-14
14	55.2	0.8	1770	4	US-09-377-155-18
15	55.2	0.8	1776	4	US-09-377-155-10
16	55.2	0.8	1776	4	US-09-377-155-20
17	55.2	0.8	1779	4	US-09-377-155-3
18	55.2	0.8	2308	4	US-09-377-155-1
19	53.6	0.8	1785	4	US-09-377-155-6
20	53.6	0.8	1785	4	US-09-377-155-8
21	53.6	0.8	1800	4	US-09-377-155-14
22	52	0.7	1779	4	US-09-377-155-16
23	48.2	0.7	15222	2	US-08-801-898A-23
24	48.2	0.7	15222	4	US-08-962-690-12
25	48.2	0.7	15223	2	US-08-892-403A-1
26	43.4	0.6	894	2	US-08-467-963C-28
27	43.4	0.6	894	2	US-08-838-189D-28

28	43.4	0.6	894	3	US-08-852-344D-28	Sequence 28, Appl
29	43.4	0.6	894	3	US-08-344-639E-28	Sequence 28, Appl
30	43.4	0.6	920	2	US-08-467-963C-7	Sequence 7, Appl
31	43.4	0.6	920	2	US-08-838-189D-7	Sequence 7, Appl
32	43.4	0.6	920	3	US-08-852-344D-7	Sequence 7, Appl
33	43.4	0.6	920	3	US-08-344-639E-7	Sequence 7, Appl
34	43.4	0.6	920	4	US-08-467-969A-7	Sequence 7, Appl
35	43.4	0.6	920	4	US-08-467-961A-7	Sequence 7, Appl
36	43.4	0.6	920	4	US-08-001-554A-7	Sequence 5, Appl
37	42.2	0.6	1236	2	US-08-741-134-5	Sequence 5, Appl
38	42	0.6	3489	2	US-08-728-323A-1	Sequence 1, Appl
39	42	0.6	32207	2	US-08-770-379-20	Sequence 20, Appl
40	42	0.6	32207	4	US-08-757-669A-20	Sequence 20, Appl
41	41.2	0.6	390	4	US-09-197-649-7	Sequence 7, Appl
42	38.4	0.6	1056	4	US-09-363-189B-3	Sequence 3, Appl
43	38	0.5	1760	1	US-08-413-118-117	Sequence 117, App
44	38	0.5	1760	3	US-08-473-446-117	Sequence 117, App
45	38	0.5	2280	1	US-08-220-151-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1  
US-08-478-370-1  
; Sequence 1, Application US/08478370  
; Patent No. 5808024  
; GENERAL INFORMATION:  
; APPLICANT: SASAKI, Ken  
; APPLICANT: HARKNESS, Robin E.  
; APPLICANT: LOOSMORE, Sheena M.  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER  
; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/478,370  
; APPLICATION NUMBER: US/08/478,370  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-502  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6973 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-478-370-1

Query Match 37.2%; Score 2580.6; DB 1; Length 6973;  
Best Local Similarity 66.9%; Pred. No. 0;  
Matches 4803; Conservative 0; Mismatches 1104; Indels 1270; Gaps 29;  
Qy 1 atgaatcacatctataaagtctcttttaacaaagccacaggttatgtcggtggca 60  
|||||

Db 542 ATGAATCACATCTATAAGTCATCTTTTAAACAAAGCCACAGGCACATTTATGGCAGTGGCA 601  
 QY 61 gaggcgcaaatccacagc-9gagggagtagoagtagtaccacagacaggtggcgag 119  
 Db 602 GAGTAGCCCAATCCACACAGCGGGGGGGTACGTGTCTACAGGGCAAGTGGCAG 661  
 QY 120 ctctctgtcaccgcctgactcgtgtgtccacgctcgtctatccttagtgcgtgcgag 179  
 Db 662 TGTATGCATCTGAGCTTTGGCCGATTTGCCGCGCTCGCTGTCTGTATGCGGTGCAAC 721  
 QY 180 gctcaatggcagtgcttatgtctaaa-----ataatagcaagatcgcatttgt- 228  
 Db 722 GCTCAGTGGCAGTGTATGCTCAAAAAAAGATACCAAAACATATCGCAATTTGGTGAACA 781  
 QY 229 ----accacaggcaaatgacaatgctcggctagcaatgaagcatcattgctattgg 284  
 Db 782 AAACAGCAGGCTGAGGCACTGCCAAGGGGAGCGGTGATCGAGCCATTTGCTATTGG 841  
 QY 285 tagtctgtcaagcacatgcgaatcaagctattgtcta cgggtggtagcaaacacagatcc 344  
 Db 842 TGAATATGCTAACGCACAGGGCGGTCAAGCCATCGCCATCGGTAGTAGTAA---TAAAC 898  
 QY 345 tcgtaatacagcggcgaatcagaagcaggttccacgcgaaggttaaaggtccatcgc 404  
 Db 899 TGTCAATCGAAGCAGTTTGGATAAGATAGGTACCGATGCTACGGGTCAAGATCCATCC 958  
 QY 405 catcgggtggtgactgactgaggtgagtgctcgtcgtgactgactgaggtgagtgactt 464  
 Db 959 CATCGGTGGTGTAAAGGCTAGTGTGTGCTGCTGATTTGCCATCGGTAGTGTGACTT 1018  
 QY 465 atatttgtagggaatagcactaacttaataatacaaatgg---tcttcttagcactct 521  
 Db 1019 ACATTTGCTGATCAGCATGATGATATCTTAACATFCCGAAGGTACTCTGATTAACGATCT 1078  
 QY 522 tattcaaacattacagattacgcgaatcagagactcaaatgggtctctc---gaata 578  
 Db 1079 TATTAAACGCCATGCAGTATTTAAAAAGAAATACGAAGTCTCAAGAGTAAATGATGATA 1138  
 QY 579 tagacgcacagcagcagaagacacgcagcagcagcagcagcagcagcagcagcagcagc 638  
 Db 1139 TAGAGGCACACACGACGCGGACGCGGACGCTGCTGAGTGGAGCCATGTCATATGCACA 1198  
 QY 639 gggctatttgcacgccttgggtacacggtcgaacagcagcagcagcagcagcagcagcagc 698  
 Db 1199 GGGTCATTTTCCACGCTTTGGTACACGCGGCAACAGCTAAAAGTGCCTATTCTTGGC 1258  
 QY 699 agtaggtcttacgcgcgaagccgaaggaatatacaatcgtcttgggttctcaatgcaca 758  
 Db 1259 AGTGGGTCTTGGCGCCACAGCGGCGCAATCTACAATCGCTATTGGTCTGTATGCAAC 1318  
 QY 759 agctaatctatggaactagccttgggtgcagatactcaggtgattggtgattcagc 818  
 Db 1319 ATCTAGCTCGTTGGGAGCGATAGCCCTTGGTGCAGGTACTGCTCAGCTACAGGCGAG 1378  
 QY 819 tattgocctagggttagtgggtctcagatccttaataataataataataataataataata 875  
 Db 1379 TATTGGCCATAGGTCAAGTCTTGTGCTACTCAGAGTGATAATAATTTAGACCGGCTA 1438  
 QY 876 agcctatgtaccagaaggaaglaattgggtcacaacataaaatcgtctaaagccacagcgaatgg 935  
 Db 1439 TACACCAAAATACCCAGGCATAGACCCCAAGTTTCAAGCCACCAATTAATACGAAGCGGG 1498  
 QY 936 ttattttccattggtagtagcactatcaagcgtaaaatcaatcaatgctcggtcaggtta 995  
 Db 1499 TCCACTTTTCCATTTGGTAGTAACTCTATCAACAGTAAATTCATCAATGTGCGGTGAGT 1558  
 QY 996 tgaagatcacatgcgtcgaatgtgcacagcgaagcgggtggagaatcctggcta---a 1052  
 Db 1559 TAATAAACCAGTCGGTCAATGTGGCACAGCTAGAAGCGGTGGTGAAGTGGGCTAAGGA 1618  
 QY 1053 gcgtcaaatctattttaaagggatgatgaacacgactgagcgtttaaagaaaaaacctggcga 1112  
 Db 1619 GCGTAGAATTAATTTTTCAGGGGTGATGATAACAGTACTGACGTAAAAAATAGGTTGGATAA 1678

QY 1113 gactttaaccattaaaggtgggtgagacccaagcgacagctaaccggataataataacat 1172  
 Db 1679 TACTTTAACTATTAAGGTGGTG---CAGAGACCAACGCATTAACCGGATAATAATATCGG 1735  
 QY 1173 tgggtggtaacagataataataactgggtctgaaagctaaactgtgtaaaaccccaagcgg 1232  
 Db 1736 TGTGGTAAAAGAGGCTGATAATAGTGGTCTGAAAGTTAACTTGTAAACATTTAAACAA 1795  
 QY 1233 tcttgaacaagtttagcacccaaacacccacccagcgagagaaagttacggtag---gtag 1289  
 Db 1796 TCTTACTGAGGTGAATACAACTACATTAATGCCACAACACAGTTAAGGTAGGTAGTAG 1855  
 QY 1290 tggtaataacacacgcgtgagctacaaagcgggtgtttbaaccttacc-----caacaac 1343  
 Db 1856 TAGTAGTACTACAGCTGAATTTATTGAGTGTATGTTTAACTTTTACCACGCCCAATACAGG 1915  
 QY 1344 aaatgcaagcaacagcaaaaaccctctatggcactgaggtcttaagtttactgataaattc 1403  
 Db 1916 CAGTCAAGCACAGCAAAACCCCTATGGCGTAAATGGGGTGAAGTTTACTAATAATATGC 1975  
 QY 1404 taatcagcgcacttga---gatactactcgtatcaccaagaataaaattggttttagcaa 1460  
 Db 1976 AGAAACACACAGCACTCGCCTACTCTGATTTACCAGAGATAAAATTTGGCTTCTCG 2035  
 QY 1461 taaagctgggtacagttgatgaacaaacaccccttcttctgataaagacaaagcctaaag--- 1516  
 Db 2036 AGATGGTGTATTGATGAAAACAAAGCACCATTATTGGATATAAANAACACTTAAAGTGGG 2095  
 QY 1517 -----t 1517  
 Db 2096 TAGTGTTCCAATTTACCATAGACAAATGCCATTTGATGAGGTAAATAAAGATCATCATCT 2155  
 QY 1518 tggcaacagcaccctaaacaacggtggtgactgacgtgtaataacacccattggtgtagcaa 1577  
 Db 2156 TGCCAAAGGTAGCAGTGTCAACGATGCGGTACCATCGACAGCTCAAGCGCCCAAGCC 2215  
 QY 1578 taaacaataacacgctgctgctgagcattaaatttgcgcgtgctgaggttaattgctatc 1637  
 Db 2216 TACTTTAAACGCGCGCTGGCCTGACAGTGTACAGCTTACTGAAATATCAGTTGATGTAA 2275  
 QY 1638 aaatgccgcgaattcggcactcgtctatcaccgaagagaaatttggcttctgctgac 1697  
 Db 2276 GAGTGGCAATTTAGCGGCCCACTTACAACATTTGGCGTGAACACCACCGAGCTTAACAG 2335  
 QY 1698 tgaatggaagttgataa----- 1715  
 Db 2336 TGATGGCACTAGTGATAAATTTAGTGTAAAGGTAGTGTACGAACAATAGCTTAGTTTAC 2395  
 QY 1716 -----aaegtcccaatatttgataaaaaaacacttcaagtggtggt 1760  
 Db 2396 CGCCGAACATTTGGCAAGCTATCTAAATGAAGTCAATCGACGGCTGACGTGCTTACA 2455  
 QY 1761 taaattaccaaagacagtggtcatttaattgaggtgatacaaaagatcagtaattgtaaga 1820  
 Db 2456 AAGCTTTTACCCTTTAAAGAGAAGACAGCATGTAGCGCAACAGCTATCACCGTGGCTAAAGA 2515  
 QY 1821 tgaacaggaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1880  
 Db 2516 TACGACAAAAAATGCGCGCGCAGTACGACATCTTAAACCTCAAAAGGTAAACCGGTCTAAC 2575  
 QY 1881 cgaagtgccctcaaa-----agcttctctattcgtgataaaaaaggttcagga 1928  
 Db 2576 GGTGTGTACCAAAAAGATGTTAGCGGTACCTTTGGCTTAGCCAAGATAGCGTCTGAC 2635  
 QY 1929 attacgatagtaactgatttcttaattgataaccccaatacctcttgagacacac 1988  
 Db 2636 CATTTGGCAAAAGCACCCTAAACACAGTGGCTTGACTTGTTTAAAGATACCAACCAACAAAT 2695  
 QY 1989 ctttgcaggtgaaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2047  
 Db 2696 CCAAGTCGGTCTAATGGCATTTAAATTTTACTAATGTGAATGGTGTAGTAATCCAGGTACTGG 2755



D	b	4865	AATGAATGTCAAAATCAGTTCATTACAAAGAACAAAGTAATGATGCCAATATAAAAGCAGG	4924
Q	y	4206	catcaatgaagacaacgcttttatcaagggcttgaaacgcgcgcaaaagacacacaaac	4265
D	b	4925	CATCAATGAAGACACGCGCTTGTGTTAAAGGACTTGAAAGACCGCTTCTGATATAACAAAC	4984
Q	y	4266	caaaaacgcgcgcagtaactgtgggtgatttaaatgcgcttcccaaaacacgcctgaactt	4325
D	b	4985	CAAAAACCGCGCAGTAACGTGTGGGTGATTTAAATGCCGTGCCCCAACACCGCTGACCTT	5044
Q	y	4326	tgacgggatacagcacaacgcgttaaaaaactggcgagactttgaccatcaaaaggtgg	4385
D	b	5045	TCAGGGGATACAGGCACACACGCGCTAANAARCTGGCGNGACTTTGACCATCAAAAGGTG	5104
Q	y	4386	gcaacagacacccaataagctaacgcgataataacatcgtgtgtgtgtagcaggtactgtatg	4445
D	b	5105	GCAAAACAGACCAATAAGCTAACCGATATAACATCGTGTGTGTAGCAGGTACTGTATGG	5164
Q	y	4446	cttcaactgtcaaaacttgcaaaagacctaaccaatcttaacagcgttaaatgcaggttggcac	4505
D	b	5165	CTTCACTGTCAAACTTGCCCAAGACCTAACCAATCTTAACAGCGTTAATGCAGGTGGCAC	5224
Q	y	4506	cgaattgatgaaaaagggcatctctttgtagacgcaaacgcgttcgaagccaaagcaaacac	4565
D	b	5225	CAAATTGATGACAAAGGCGTGTCTTTGTAGACTCAAGCGGTCAAGCCCAAGCAAAACAC	5284
Q	y	4566	ccctgtctaaagtgcgaatgggctggacctgggtggcaaacgcatacagtaacatcgggtgc	4625
D	b	5285	CCCTGTGCTAAAGTGCCAAATGGGCTGGACCTGGGTGGC-----	5321
Q	y	4626	agcgtgtgatgaalagcgtgcggtagaactttaagcagtttaagtaagtgtgcacaaacgggt	4685
D	b	5322	-----	5321
Q	y	4686	caacaacctaaacaacaaaagtaactcaggtgcgtcattaccctttgttgtaacogatgc	4745
D	b	5322	-----	5321
Q	y	4746	caatggcaagcccatcaatggcaccgatggcaagcccccaaaagccatcaaggggcgca	4805
D	b	5322	-----	5321
Q	y	4806	tggtaatactatacgcgcaaacccaaacgcgttacctgttggaacaaagatggcaagcccat	4865
D	b	5322	-----	5321
Q	y	4866	caccgatgcggacaaaacttgccaatctggcagctcatggaaccccccttgatgcaggtca	4925
D	b	5322	-----	5321
Q	y	4926	tcaagtgtggcaagcctagcggcaactcagatgcatacccttaacccaacatcaagtc	4985
D	b	5322	-----	5321
Q	y	4986	cactttgccacaattgacacaccaaacaacaggttaatgccaatgcaggggcaagcccaag	5045
D	b	5322	-----	5321
Q	y	5046	tctgccagcctatacagcagcagcaagaatgtgtgcagtgtaaaagatgtgctaaa	5105
D	b	5322	-----	5321
Q	y	5106	tgtaggctttaacttgagaccatacaatacaagtggactttgtcaaaagcctatgatac	5165
D	b	5322	-----	5321
Q	y	5166	cgtcaactttgtcaatggtacaggtgcgcagatcacaaagcgtgcgtagtgatggcac	5225
D	b	5322	-----	5321
Q	y	5226	gatgagtaacatcacccgtcaacacgcgcttagcagcgacogatgatgatggcaatgtgct	5285
D	b	5322	-----	5321

Q	y	5286	tatcaaaagccaaagatgggtaagtcttacaagaagcagacgacctcatgccaaaacgcgtcact	5345
D	b	5322	-----	5321
Q	y	5346	aaaagcaggcaaatcagccagtgatgccaaaactccaaactggctaaagccttgttaacccc	5405
D	b	5322	-----	5321
Q	y	5406	caatgctgggtaaaggcagtagcagcgatgcagtggtccttaataaacttatcaaaaagcggt	5465
D	b	5322	-----	5321
Q	y	5466	atttaaatccaaagatggtatacactactaccagtaagctctgatggcatcagtatccca	5525
D	b	5322	-----	5321
Q	y	5526	aggcaaaagataaacagcagcatcacccctaagcaaaagatgggctgaatgtaggcggtgaagt	5585
D	b	5322	-----	5326
Q	y	5586	catcagcaatgtgggttaaaggcacaagaagacacgcgacgtgcgcaatgtacaaacagttaaa	5645
D	b	5327	CATCAGTAAATGTGGCAAGGCACAAAGATACCGACGCTGCCAATGTACAAACAGTTAAA	5386
Q	y	5646	cgaagtacgcgaactgtttgggtccttggttaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	5705
D	b	5387	GAAAGTACGCCAACTTGTGGGTCTTGGTAAATGCTGGTAATGATAACGCTGACGGCAATCA	5446
Q	y	5706	ggtaaacattccacatcaaaaaagacccaaatctcaggttctcatcttaaccgcactgt	5765
D	b	5447	GGTAACATTCGCCGACATCAAAAAGACCCAAATTCAGGTTTCATCTAATCCGCACACTGT	5506
Q	y	5766	catcaaaagcagcaggttacttggcgggttaaagggttaaataacagatacccgaaaaacttgcac	5825
D	b	5507	CATCAAAAGCAGCAGCGGTACTTGGCGGTAAAGGTAAATACGATACCGAAAAACTTGCAC	5566
Q	y	5826	tgggtgtgtacaagtggcgtggataaaagacgcgaacgtaacgctgaacgctttaaagcaatgt	5885
D	b	5567	TGGTGGTATACAAGTGGCGGTGGATAAAGCGCAACCGCTAACGGCGATTTAAAGCAATGT	5626
Q	y	5886	ttgggtcaaaaaccccaaaagatggcagcaaaaaagccctgtctgcaccttaaacgcgcgc	5945
D	b	5627	TTGGGTCAAAACCCCAAAAGATGGCAGCAAAAAGCCCTGTCTGCCACTTATTAACGCCGC	5686
Q	y	5946	aggtcagacccaactatttgacaacaaacccgcagaaagcattacagataaaatgaaca	6005
D	b	5687	AGGTGAGACCAACTATTTGACCAACACCCCGCAGAGCCATTGACAGATTAATATGAACA	5746
Q	y	6006	aggtatccgcttcttccatgtcaacgattggcaatcaaaagcctgtgggtacaaagggcgttaa	6065
D	b	5747	AGGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCCCTGTGGTACAAAGGCGCTAA	5806
Q	y	6066	cggcattgactcaagtgcctcaggaagcactcagtggcgataagtttccaaagccaaagcc	6125
D	b	5807	CGGCATTGACTCAAAGTGCCTCAGGCAAGCACTCAGTGGCGATAGGTTCAGCGCCAAAGCG	5866
Q	y	6126	agatgtgaagccgcgcttggccataggcagacaaaccccaagcagcgaacaaactccatgc	6185
D	b	5867	AGATGTGAAGCCGCCGCTTGCCATAGGCAGACAAACCCCAAGCAGGCAACCAATCATCGC	5926
Q	y	6186	catcgtgatgaacgcacaagccacggcgatcaatccatcgcacatcgggtacagggcaatgt	6245
D	b	5927	CATCGGTGATACGCACAAGCCACGGCGCATCAATTCATTCGCCATTCGGTACAGGCAATGT	5986
Q	y	6246	ggtaaacaggtaaagcacacttgggtccatcgcgcaccccaagcactgttaagcctgataaacg	6305
D	b	5987	GGTAGCAGGTAAAGCACTCTGGTGCCATCGCGACCCCAAGCACTGTTAAGGCTGATTAACAG	6046
Q	y	6306	ttacagtggtgtgttaataacaacacagtttatcagatgcacatcagacccgagtgtctttggtgt	6365
D	b	6047	TTACAGTGTGGGTAAATAACAACACCGTTTACCGATGCCATCAAAACCGATGCTCTTTGGTGT	6106

```
QY 6366 gggcaataacatcacccgtgaccggaagtaactcgttgctttaggttcaaacctctgccat 6425
Db 6107 GGGCAATAACATCACCGTGACCAGAAAGTAACTCGGTGCTTGGTTCAAACTCTGCCAT 6166
QY 6426 cagtgcagcacacacgacgacacaaagcaaaatctgacgacacacaggtacaac 6485
Db 6167 CAGTGACGGCACACACGACGGCACAAAGCCAAATACTGACGGCACACAGGTACAAC 6326
QY 6486 caccacagcaggtgcaacaggttacggttaaaggcttttctggacaaacgcggttggtgc 6545
Db 6227 CACCACAGCAGGTGCAACCGTACGGTTAAAGGCTTTGCTGGACAAACGCGGTTGGTGC 6286
QY 6546 ggtctcgtgggtcgtcaggtgctgaaccccgctatccaaaaatcgtggcagcaggttaggt 6605
Db 6287 GGTCTCGTGGGTGCTCAGGTGCTGAACCGCTATCCAAAATGTGGCAGCAGGTAGGT 6346
QY 6606 cagtgcaccacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 6665
Db 6347 CAGTGCCACACGACCGCATCGGTCAATGTTAGCCAGTTGTAACAAAGCCACCCAAAGCAT 6406
QY 6666 tgccacgacacacaaatgagcttgaccatcgtatccaccacacacacacacacacacacac 6725
Db 6407 TGCCACGCAACCAATGAGCTTGACCATCTATCCACCAAAACGAAAATAAGGCCAATGC 6466
QY 6726 agggatttcacgagtgatgagtgatgagtgatgagtgatgagtgatgagtgatgagtgatg 6785
Db 6467 AGGGATTTCATCAGCGATGCGATGCGCTCCATGCCACAAAGCCCTACATTCCTGGCAGATC 6526
QY 6786 catggttacgagggtattgccaccacacacacacacacacacacacacacacacacacac 6845
Db 6527 CATGTTACCGGGGGTATGCCACCCACACCGTCAAGGTGCGGTGGCAGTGGACATGTC 6586
QY 6846 gaagctgcgataaatggtcaatgggtattttaaatacaatgggttcacgacgacacacacac 6905
Db 6587 GAAGCTGCGGATAATGCTCAATGGGTATTTAAATCAATGGTTTCAGCCGATACCCCAAGG 6646
QY 6906 ccatgtaggggcgagcttggtgagcttttcacttt 6942
Db 6647 CCATGTAGGGGCGCAGTTGGTGCAGGTTTTTCACTTT 6683
```

```
RESULT 2
US-08-968-685A-9
; Sequence 9, Application US/08968685A
; Patent No. 6214981
; GENERAL INFORMATION:
; APPLICANT: TUCKER, KENNETH
; APPLICANT: PLOSILA, LAURA
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968, 685A
; FILING DATE: No. 6214981ember 12, 1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7969-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
```

```
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-968-685A-9

Query Match 34.3%; Score 2382.4; DB 4; Length 9542;
Best Local Similarity 65.2%; Pred. No. 0;
Matches 4842; Conservative 0; Mismatches 1041; Indels 1545; Gaps 29;

QY 1 atgaatcacatctataaagtcattctttaacaagaagccacagccacatttatggcgtggca 60
Db 218 ATGAATCACATCTATAAAGTCATCTTTTAAACAAAGCCACAGGCACATTTATGGCCGTGGCG 277
QY 61 gagtgcgcacaaatcccccacagcgaggagtagcagtagtaccgcagcagcaggtgggcagc 120
Db 278 GAATATGCCAAATCCACACACGCGGGGGGTAGCTGTCTACAGGCAAGTTGGCAGT 337
QY 121 tctctgtcatccgcgtgactcgtgttgccacgctcgtctatccttagtgcagtgagcag 180
Db 338 GTACGCACTCTGAGCTTTGCCCGCTATTTCGCGCGCTCGCTGCTCGTATCGGTGCGACG 397
QY 181 ctcaatggcagctcttatgctcaaaaataatagaagaatcgactttgtaccacaggaac 240
Db 398 CTCAATGGCAGTGTCTATGTCAGGTATCGGAATTAGTGAAGCAGACGGGGGAAAAAGG---- 453
QY 241 aatgacaatgcctcggttagcgaatgaagcatccattgttcttagtcttctgctlaaggca 300
Db 454 ----CGAGCCCAATCGCGGGTGATAAATCCATTGCTATTGCTGATATTGCTCAGGCA 508
QY 301 catgccaatcaagctattgctatcgtgtggttagcgaacacagatcctctaatcaagcgct 360
Db 509 CTTGGCTCTCAATCTATTGCTATCGGTGACAAACAAAATAG---TTCAATAATCAATAAT 565
QY 361 aalcagaaggcaggttcccacgcccacaaaggtaaaagagtcacatcgccatcggttgatgta 420
Db 566 AATGCTAATATAGGTGCCAAGCCTCAGGTCAAGTAATAGTCCATCGCCATCGGTGGTGATGTA 625
QY 421 ctggctgagggtagctcctcgattgcttagtgatgacttatttggataggagaat 480
Db 626 TTGGCTTCTGGTCATGCTCGATTGCCATCGGTAGTGTGATGACTTATATTGAAAAAGGA- 684
QY 481 agcactaaccttaaatatccaaatggtctcttagcactcttattcaaaacattacagta 540
Db 685 --AACGGTACAGCAAAATCTCAGAGCTTCTACCTATTATTTCGCGGACAGAAAGCATTAAC 742
QY 541 ttacgccaatacagagactcaaatggtttctcagaataatagacacagcagcagagagga 600
Db 743 GATATATACCAACTAGCTGACACTAAATCTTCAAAAAATATAGACGCACACAGCAGGGA 802
QY 601 cagccagtagctcagtgaggagccatgcatatgcaagggttcatttgcacacgccttt 660
Db 803 CAGCCAGTACTCAGTGGGAGCCATGCTATATGCAAGGGTCATTTTCCAAAGCCTTT 862
QY 661 ggtacaggtcaacagctgaagcgaactattccttggcagtaggtctctaccgcgaagcc 720
Db 863 GGTACACGGGCACACAGCTGAAGGTACCTATTTCCTTGGCAGTGGGTCTTACGCCACAGCC 922
QY 721 gaaaaagatatcaaatcgctattggttctaaatgacacaaagctatcaattatggagacta 780
Db 923 AAAGCAGCATCTTCAATCGCTGTTGGTTCTAATGCAAGCTATCGGGTTTGGACGACA 982
QY 781 gcccttggtgcagatactcagttgatttgattacaggtattgccttaggttatggttct 840
Db 983 GCCCTTGGTGAAGTACTCAAGTTAAATTTGAATCGAGGTATTGCCCTAGGTTTTGGTTCT 1042
QY 841 cagatccttaataataataataataataataa-----gcctatgtaccagaa 891
```

[illegible]



QY 2599 gaaacttaaacacccctagccaaggaattcaacaccccaaaagggcacagcagacacagcc 2658  
 Db 3242 GACAATCTAAACACCCCTAGCTGTGTAATTCGCAACACCAAAAGGCACAGCAGACCCGCC 3301  
 QY 2659 ctcaaacaccttaaaagtcacaaaa-----gacggtcaactgatgacgaa 2703  
 Db 3302 CTACAAACCTTTNMGTCANAAAGTCAAGANAAATGGTGATGATATAGCGCTGAC 3361  
 QY 2704 accatccctgggttaagatggtcacaaaaagcgcaagccgtcaacactctaaactc 2763  
 Db 3362 ACCATCACCGCTGGGTAAGATG-----CAAAAACCAATCAAGTCAACACCCTAACACTC 3415  
 QY 2764 aaagtgaaacgggtctaaacgggttgctaccaaataaagatggttacggttacctttggcatt 2823  
 Db 3416 AAAGGTAAACCGCTCTTGATATTCAAAACCAATAAAGATGCTACGGTTACCTTTGGCATTT 3475  
 QY 2824 aacacccaagcggtcttaagcgcgagcagcaccactctaaacaaagatggcttctct 2883  
 Db 3476 AACACCCANAGCGGCTTAAGCGGGCAACACACCACTCTNACACANATGGCTTGCT 3535  
 QY 2884 attaaaccccgctagtaacgaacaaatccaagtcggtgctgatggcgtaagttggcc 2943  
 Db 3536 ATTAACCAACCGCTGCTACGAAACAAATCCAAGTCGGTCTGATGCGTGAAGTTTGGC 3595  
 QY 2944 aaggtgataaagggttaattcaagcactggcattgatggcacaagccgtatcaacaaagat 3003  
 Db 3596 AAGGTTAATAATGGTGTGTAGGTGCTGGCATTTGATGCCACAACTCGCATTTACAGAGAT 3655  
 QY 3004 caaattgctttactgggttaagtcactgctacttgatcaccacaaaccccaactaaacaaa 3063  
 Db 3656 GAAATGGCTTTGCTGGGACTAATGGCTCACCTTGATAAAAGCAAAACCCCACTTAAGCAAA 3715  
 QY 3064 gacaagcttaagtggtgaagttgaattaccaacactggcattaaacgaggtgggttaa 3123  
 Db 3716 GAC-----GGCATTAACGCGAGGTGGTAAA 3739  
 QY 3124 aagattaccaacttaactcaaggtgatattaccacaaacagcgaatgatgctgtaacggc 3183  
 Db 3740 AAGATTACCAACATTCATCAGGTGAGATTGCCCAAAACAGCAATGATGCTGTGACAGGC 3799  
 QY 3184 ggtcggtttatgatttaaaacccgaacttgaagcaaaatcaacagtgctgctaaacaa 3243  
 Db 3800 GGCAAGATTATGATTAAACACCGCACTTGAAACACAAATACAGCACTACTGCCAAACAA 3859  
 QY 3244 gcacaaactcaatcacgaattctcagtagcagatgaacaaaggttaactcatttaccggtt 3303  
 Db 3860 GCACAAAACCTATTACACGAATTTCTAGTAGCAGATGAACAAGGTAATACTTTACGGTT 3919  
 QY 3304 agtaaccttactccagttatgacacctcaagacctctgatgctacacctttgacaggt 3363  
 Db 3920 AGTAACCCCTTACTCCAGTTATGACACCTCAAGACCTCTGATGTGCATCACCCTTTGACAGGT 3979  
 QY 3364 gaaacggcattaccaccaaaggttaataaaggttgctgctggtggcattgaccaaacc 3423  
 Db 3980 GAAACGGCATTACCAACCAAGGTAAATAAGGTGTGCTGTGGCATTGACCAAAACC 4039  
 QY 3424 aaaggttaacacgcgttaagctgacggtgggttaataatggtcaaaagcgattgtcatt 3483  
 Db 4040 AAAGGCTTAACCAACCGCTTAAGCTGACCGTGGGTAAATAATATGSCAAAGGCATTGTCAAT 4099  
 QY 3484 gacagtaaaagtgtcaaaataaccacaggaactaagcaacactctagtaattgttacc 3543  
 Db 4100 GACAGCCAAAATGGTCAAAAATPACCATCACAGGACTAAGCAACACTAGCTAATGTTTACC 4159  
 QY 3544 aatga-----tggtgcagacacgcactaagccaagggcttgccaatgacacgcacaaa 3597  
 Db 4160 AATGATAAGGTACCGTACCGCACCACAGACAGAGCGGCAAGATATCAAGAGCAAGACAAA 4219  
 QY 3598 acccggtccgcagcattggtgatgtgctaaacgcagcgcttttaacttgcaagggaattggt 3657  
 Db 4220 ACCCGTCCGCCACGATTTGTTGATGTGCTAAGCGCAGGCTTTAACTTGCANAGGCAATGGT 4279

QY 3658 gaagcggtgactttgctctccacttatgacactgttgactttatcgatgccaatgcaacc 3717  
 Db 4280 GAAGCGGTGACTTTGCTCCACTTATGACACTTATGACACTTATGCGGCAATGCGCAACC 4339  
 QY 3718 accgctaagtgacttatgatgacaaagcaaaacacagtaagtggtctatgatgtcaat 3777  
 Db 4340 ACCGTAAGGTGACCTATGATGACAAAGCAANAAACCAAGTAAAGTGGTCTATGATGTCAAT 4399  
 QY 3778 gtggataaaaaacattgaagtgcacagtgataaaaaacttggctcaaaacaccacaca 3837  
 Db 4400 GTGGATCATCAACCACTTGAAGTTAAA-----GATAAAAAAACTTGGCGTAAAAACACCACCA 4456  
 QY 3838 ctgacaaaaacaagtgttaattgtaagtca-----accaaatttagtgcgcgcagatggcgat 3894  
 Db 4457 TTGACCAAGTACTGGCAGAGTGTCTAATAAATTTGCCCTAAGCAATCAAGTACTTGGCGAT 4516  
 QY 3895 gcccttgaagccagtgatcgccaccatctaaactctggctggcgacatccaa 3954  
 Db 4517 GCGTGTCTAAGCGCAGTGATATGCTGCTCTCTTAACACCTTATCTGGCGACATCCAA 4576  
 QY 3955 accgccaaggggcaagcccaagcagctcagcaagctatgtggatgctgtatggcaac 4014  
 Db 4577 ACTGCCAAAGGGCAAGCAAGCAACAGCTCAGCAGGCTATGTGGATGCTGATGGCAAT 4636  
 QY 4015 aaggtcacttatgacagtcacagataaagaagtactatcaagtcaatgacaaaggttcaagtg 4074  
 Db 4637 AAGGTCTATCTATGACGTACCGGATTAACAGTACTTATCAAGCCAAAAATGATGGCACATTT 4696  
 QY 4075 gacaaaaacaaagagtgccaaagacaaactggtcccaagcccaaacccacagatggc 4134  
 Db 4697 GATAAAACCAAGAGTTGCCAAAGCAAACTGCTGCCCAAGCCCAACCCCAAGATGGC 4756  
 QY 4135 acattggtctcaaatgaatgcataatcagtcattacaaaagagcaagtaaatgatgccaat 4194  
 Db 4757 ACATTGGCTCAATGAATGTCAATCAGTCATTAAACAAAGAAACAAGTAATGATGCCAAT 4816  
 QY 4195 aaaaagcaagcactcaatgaagacaaacgctttatcaaaagggcttgaacacgcgcgcaaa 4254  
 Db 4817 AAAAAGCAAGGCATCAATGAAGACCAACGCCCTTTGTTAAAGGACTTGA AAAAGCGCTTCT 4876  
 QY 4255 gacacccaaacccaaacgcgcagtaactggtggtatttaaatgcccgttgcacaaaca 4314  
 Db 4877 GATACAAAACCAAAAACGCGCGAGTAACTGTGGGTGATTTAATGCGTTGCCCAACAA 4936  
 QY 4315 ccgctgacctttgaggggatacagggcacaaacgcgcgttaaaactggcgagactttgacc 4374  
 Db 4937 CCGCTGACCTTTGACGGGATACAGGCACAACGGCTAAAAAATCTGGCGACACTTTGACC 4996  
 QY 4375 atcaaaagtgggcaaaacagacaccaaataagctaaaccgataataacatcggtgtgtagca 4434  
 Db 4997 ATCAAGGTGGGCAAAACAGACACCAATTAAGCTAACCCGATAATAACATCGGTGTGGTAGCA 5056  
 QY 4435 ggtactgaggttcaactgtcaaaactgcaaaagacccctaaacactcttaacagcgttaat 4494  
 Db 5057 GGTACTGATGCTTCACTGTCAAACTTGCCAAAGACCTTAACCAATCTTTACAGGCTTAAT 5116  
 QY 4495 gcaggtggcacaagaattgatgaaaaagggcatctcttttttagacgcgaacacgggtcaagcc 4554  
 Db 5117 GCAGGTGGTACCAAAATTGATGACAAAGGCGCTGCTTTTGTAGACTCAAGCGGTCAAGCC 5176  
 QY 4555 aaagcaaacacccctgtgctaaagtgccaaatgggctggaacctgggtggcaaacgcatacagt 4614  
 Db 5177 AAAGCAAAACCCCTGTGCTTAAGTCCCAATGGGTGGACCTTGGGTGGC----- 5224  
 QY 4615 aacatcggtgcagctgttgatgataacgatcggtgaactttaagcagtttaataagaatt 4674  
 Db 5225 ----- 5224  
 QY 4675 gccaaaaaggtcaaacacacttaaacacaaagtaactcaggtgcgtcattaccctttgtg 4734  
 Db 5225 ----- 5224  
 QY 4735 gtaaccgatgccaatgccaagcccccaatggtgacccgatggaagcccccaaaagccatc 4794



	Query Match	1.28;	Score 83.6;	DB 4;	Length 2037;
	Best Local Similarity	54.1%;	Pred. No. 3.4e-14;		
	Matches 196;	Conservative	0;	Mismatches 159;	Indels 9; Gaps 1;
Qy	6584	aaatgtgcagcagggtgagctcagtcgccaccagcacagatgcggtcaatagttagccaagt	6643		
Db	1670	ANAAAGTCGCGGCGAGGGGAAATTTCTGCCACTTCCACCGATGCGATTAAACGGAAGCCAGT	1729		
Qy	6644	tgtcaaaagccaccacagggcattggccaaagcaacc-----aatgagctttgaccatc	6694		
Db	1730	TGTATGTCGCGGCGAAAGGGGTAAACAAACCTTGCTGGCAAGTGAATCTTGAGGCCA	1789		
Qy	6595	gtatccaccacaaacgaaaaataaagccatgcagggatttcatacgcgaglgcgatg9cgt	6754		
Db	1790	AAGTCAATATAAGTTGGGCAAAACGTCGAGATGAGCTACTGTCAAGTGCATTTAGCGGGTTCAC	1849		
Qy	6755	ccatgccacaagcctacattctcggcagatccatggttaccg9gggtattgccaccaca	6814		
Db	1950	AGTTACCACAGGCCACATATGCCAGGTAATCAATGGTTTCTATTTCGGGGAAGTAGTTATC	1909		
Qy	6815	acggtcaaggtgcggtggcagtg9gactgtcgaaagctgtcggataaatggtccaatggggtat	6874		

	Query Match	1.28;	Score 82.8;	DB 1;	Length 3294;
	Best Local Similarity	49.8%;	Pred. No. 8.4e-14;		
	Matches 244;	Conservative 0;	Mismatches 237;	Indels 9;	Gaps 1;
QY	6460	aaatctgacgcagcagcaggtacaaacaccacacacgaggtgcaacaggtgcacggtttaaagcc	6519		
Db	2803	AAAGCTGACGGTACTGCGGATAAAACCAACGCGAAAGTGAGCAATGATAAAGTTTCTACC	2862		
QY	6520	tttctggagcaaacgcggttggtgcggtctccctggtgctcaggtgcgtgaaagccggt	6579		
Db	2863	GATGAAAAAACACGTTGTGACGCTTGATCCAAATGATCAATCAAAAGGTAAAGGTGTCGTG	2922		
QY	6580	atccaaaattggcagcagcaggtgaggtcagtgccaccagcacggtcaggtcaatggttagc	6639		
Db	2923	ATTGACAATGTGGCTAATGGCGATATTTCGCCACTTCCACCGATGCCATTAAACGGAACT	2982		
QY	6640	cagttgtacaagccaccaggcattgccaacgcaccc-----aatgagcttgcac	6690		
Db	2983	CAGTTGTATGCTGTGGCAAAAGGGGTAAACAACCTTGCTGGACAAGTGAATTAATCTTGAG	3042		

Ddb	2803	AAAGCTGACGGTACTGCGGATAAAACAAAGCGAAGTGAAGCAATGATAAAGTTTCTACC	2862
Qy	6520	tttgctggacaacacgcggttgctgcggctcctcgtaggtgcctcagggtgctgaacgcgt	6579
Ddb	2863	GATGAAACACAGTTGTACGCCTTGATCCAAATGATCAATCAAAAGGTGAAGGTGTCGTG	2922
Qy	6580	atccaaaatgtggcagcaggttgaggtcagtgccacacgacacgcgattcggtcgaatgtagc	6639
Ddb	2923	ATTGACAAATGTGGCTAATGGCGATATTCTTGCACATCTCCACGATCGAATTACGGAAATG	2982
Qy	6640	caagttgataaagccaccacacgcatgtgccaaacgcaacc-----aatgagttgac	6690
Ddb	2983	CAGTTGTATGTGTGGCGAAACGGGTACAAACCTTGCTGCACAACTGAATAATCTTGAG	3042
Qy	6691	catcgtatccaccaaaaacgaaaaataagccaatgcaggaatttcacgcgattgcccgtg	6750
Ddb	3043	GGCAAAAGTGAATATAAGTGGGCANAACGTCAGATGCAGGTACAGCAAGTCATAGCGCCT	3102
Qy	6751	gggtccatgccacaagcctcatctcctggcagatccatgtgtaccggggggtatgccaacc	6810
Ddb	3103	TCACAGTTTACCACAGCCACATATGCCAGGTAAATCAATGTTGTATTTCGGGAAGTAGT	3162
Qy	6811	cacaaacggttcaaggtgcggtggcagtgccgaagcgtgcgaagcgtgcggaataatggtccaatgg	6870
Ddb	3163	TATCAAGGTCAAATGGTTTGTAGCTATCGGGGTATCAAGAAATTCGCATTAATGGCAAGTG	3222
Qy	6871	gtatttaaatcaatggttcagccgatccccaaaggccatgtagggggcgagctgtgtgaca	6930
Ddb	3223	ATTATTCGGTTGTACGGCACACCAATATGTAAGGTAAACAGCGCTTGTCACGAGGTGTT	3282
Qy	6931	ggttttcact	6940
Ddb	3283	GGTTACCAGT	3292

RESULT 6  
US-08-913-942-1  
; Sequence 1, Application US/08913942  
; Patent No. 6200578  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme, Joseph  
; APPLICANT: Barenkamp, Stephen J.  
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr Hohlbach Test Albritton & Herbert LLP  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,942  
; FILING DATE: 29-DEC-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/409,995  
; FILING DATE: 24-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/4031  
; FILING DATE: 22-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Vance, Dolly A.  
; REGISTRATION NUMBER: 39,054  
; REFERENCE/DOCKET NUMBER: A-61053-1/RTT/RMS/DAV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249

[illegible]

STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,467  
FILING DATE: 22-JUL-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/409,995  
FILING DATE: 24-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3294 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-685-467-1

Query Match 1.2%; Score 82.8; DB 3; Length 3294;  
Best Local Similarity 49.8%; Pred. No. 8, 4e-14;  
Matches 244; Conservative 0; Mismatches 237; Indels 9; Gaps

6460 aaatctgcgcgcacagcaggtgtacacccacacacagcaggtgcacacaggttcaggttaaacgc 6519

```

;
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3294 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-08-913-942-1

Query Match      1.2%   Score 82.8; DB 4; Length 3294;
Best Local Similarity 49.8%; Pred. No. 8.4e-14;
Matches 244; Conservative 0; Mismatches 237; Indels 9; Gaps 1;

Qy 6460 aaatctgacgcagcaggttacaccaccacacagcaggtgcaacagggtcacggtttaaaagcc 6519
Db 2803 AAAGCTGACGGTACTCGGTAAACCAACCAAGGCCGAAGTGACCAATGATTAAGTTTCACC 2862

Qy 6520 ttctgtggaacaacgcggttggtcggtgtcccttggtgacctcaggtgctgtaaacgcgt 6579
Db 2863 GATGAAAACACAGTGTGCAGCCTTTGATCCAAATCATCAATCAAAGGTAAAGGTGTCGTG 2922

Qy 6580 atccaaaatggtgagcaggtgaggtcagtgccaccagcaccgatcggtgcaatggtagc 6639
Db 2923 ATTGACAATGCGTAAATGGCGATATTCTGCCACTTCCACCGATGCGATTAAACGGAAGT 2982

Qy 6640 cagtgtacaagaaccaccaagcatgccaaacgaacc-----aatgagcttgac 6690
Db 2983 CAGTTGTATGCTGTGGCAAAAGGGGTACAAACCTTCTGGACAAAGTGAATAAATCTTGAG 3042

Qy 6691 catcgtatccacaaaaacgaaataaaagccaatgcagggatttcatcagcagatgcccgtg 6750
Db 3043 GGCAAGTGAATAAGTGGGCAACAGTGCAGATGCAGTACAGCAAGTGCAATTAGCGGCT 3102

Qy 6751 gcgtccatgccaagaactacattcctggcagatccatggttacgcgggggtatggccacc 6810
Db 3103 TCACAGTTTACCACAAGCACACTATCCAGGTAAATCAATGATGTTGCTATTGCGGGGAAGTAGT 3162

Qy 6811 cacaaagtcgaagtgcggtggtgagctgcgaagctgcggaataatggccaatgg 6870
Db 3163 TATCAAGTCAAAATGGTTAGCTATCGGGTATCAAGAATTCGGAATATGGCAAGTAGT 3222

Qy 6871 gtatttaaaatcaatggttcagccgatacccaaagcccatgtagggggccccagttggtgca 6930
Db 3223 ATTTATCGTTGTCAGGCACAAACCAATAGTCAAGGTAAAAACAGCGGTGTCAGCAGGTGTT 3282

Qy 6931 ggttttcaact 6940
Db 3283 GGTACCAGT 3292

RESULT 7
US-08-913-942-3
; Sequence 3, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph
; APPLICANT: Barenkamp, Stephen I.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```



COUNTRY: Canada  
 ZIP: M5G 1R7  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/478,370  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Stewart, Michael I  
 REGISTRATION NUMBER: 24,973  
 REFERENCE/DOCKET NUMBER: 1038-502  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (416) 595-1155  
 TELEFAX: (416) 595-1163  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 60 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-478-370-4

Query Match 0.8%; Score 58.4; DB 1; Length 60;  
 Best Local Similarity 98.3%; Pred. No. 6.2e-08;  
 Matches 59; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4150 aatgtcaaatcagtcattacaagaagcaagtaaatgatgccataaaagcaaggcattc 4209  
 |||||  
 Db 1 AATGTCAATCAGTCATTACAAGAAGCAAGTAATGATGCCAATAAAGCAAGGCATC 60

RESULT 11

US-09-377-155-12  
 ; Sequence 12, Application US/09377155  
 ; Patent No. 6197312  
 ; GENERAL INFORMATION:  
 ; APPLICANT: PEAK, Ian Richard Anselm  
 ; APPLICANT: JENNINGS, Michael Paul  
 ; APPLICANT: MOXON, E. Richard  
 ; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN  
 ; FILE REFERENCE: 065064/0128  
 ; CURRENT APPLICATION NUMBER: US/09/377,155  
 ; CURRENT FILING DATE: 1999-08-19  
 ; PRIOR APPLICATION NUMBER: PCT/AU98/01031  
 ; PRIOR FILING DATE: 1998-12-14  
 ; PRIOR APPLICATION NUMBER: GB 9726398.2  
 ; PRIOR FILING DATE: 1997-12-12  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 12  
 ; LENGTH: 1797  
 ; TYPE: DNA  
 ; ORGANISM: Neisseria meningitidis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(1794)  
 US-09-377-155-12

Query Match 0.8%; Score 58.4; DB 4; Length 1797;  
 Best Local Similarity 52.0%; Pred. No. 8.7e-07;  
 Matches 131; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 6689 accatcgtatccaccacaaacgaaataaaagccaatgcagggtatttcacgagcgatggcga 6748  
 |||||  
 Db 1541 acaaccgcatcgacaatgtggacggcaacgcgcgcgggtatcgcccaagcgattgcaa 1600  
 QY 6749 tggcgtccatgccacaagcctacattcctggcagatccattcgcgcgggtacgttaccgggggtattgccca 6808

Db 1601 ccgcaggtttggctcagcgctatttgcgcggaagagatgatgcgatcgcgcgcggtta 1660  
 QY 6809 ccacaaacggtcaaggtgcgggtgcagtggaactgtcgaagctgtcgaataatgtcaat 6868  
 Db 1661 ctatcgcggcgaagccggttacgcctacgcctactcgcgacatttctgacactgggaatt 1720  
 QY 6869 ggggtatttaaaatcaatggttcagcccgatcccccaagcccatgtaggcgcgagttgggtg 6928  
 |||||  
 Db 1721 ggggtatcaagggcagcgcttcgcggcaattcgcgcgcccatttcggtgcttccgcactcg 1780  
 QY 6929 cagggttttcaact 6940  
 |||||  
 Db 1781 tcgggttatcagt 1792

RESULT 12

US-09-377-155-4  
 ; Sequence 4, Application US/09377155  
 ; Patent No. 6197312  
 ; GENERAL INFORMATION:  
 ; APPLICANT: PEAK, Ian Richard Anselm  
 ; APPLICANT: JENNINGS, Michael Paul  
 ; APPLICANT: MOXON, E. Richard  
 ; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN  
 ; FILE REFERENCE: 065064/0128  
 ; CURRENT APPLICATION NUMBER: US/09/377,155  
 ; CURRENT FILING DATE: 1999-08-19  
 ; PRIOR APPLICATION NUMBER: PCT/AU98/01031  
 ; PRIOR FILING DATE: 1998-12-14  
 ; PRIOR APPLICATION NUMBER: GB 9726398.2  
 ; PRIOR FILING DATE: 1997-12-12  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1797  
 ; TYPE: DNA  
 ; ORGANISM: Neisseria meningitidis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(1794)  
 US-09-377-155-4

Query Match 0.8%; Score 56.8; DB 4; Length 1797;  
 Best Local Similarity 51.6%; Pred. No. 2.6e-06;  
 Matches 130; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 6689 accatcgtatccaccacaaacgaaataaaagccaatgcagggtatttcacgagcgatggcga 6748  
 |||||  
 Db 1541 acaaccgcatcgacaatgtggacggcaacgcgcgcgggtatcgcccaagcgattgcaa 1600  
 QY 6749 tggcgtccatgccacaagcctacattcctggcagatccattcgcgcgggtacgttaccgggggtattgccca 6808  
 |||||  
 Db 1601 ccgcaggtttggctcagcgctatttgcgcggaagagatgatgcgatcgcgcgcggtta 1660  
 QY 6809 ccacaaacggtcaaggtgcgggtgcagtggaactgtcgaagctgtcgaataatgttcaat 6868  
 |||||  
 Db 1661 ctatcgcggcgaagccggttacgcctacgcctactcgcgacatttctgacactgggaatt 1720  
 QY 6869 ggggtatttaaaatcaatggttcagcccgatcccccaagcccatgtaggcgcgagttgggtg 6928  
 |||||  
 Db 1721 ggggtatcaagggcagcgcttcgcggcaattcgcgcgggtcatttcggttacttccgcactcg 1780  
 QY 6929 cagggttttcaact 6940  
 |||||  
 Db 1781 tcgggttatcagt 1792

RESULT 13

US-08-232-463-14/c  
 ; Sequence 14, Application US/08232463  
 ; Patent No. 5670367



[illegible]

; PRIOR APPLICATION NUMBER: GB 9726398.2  
 ; PRIOR FILING DATE: 1997-12-12  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 10  
 ; LENGTH: 1776  
 ; TYPE: DNA  
 ; ORGANISM: Neisseria meningitidis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(1773)  
 US-09-377-155-10

Query Match 0.8%; Score 55.2; DB 4; Length 1776;  
 Best Local Similarity 51.2%; Pred. No. 7.6e-06;  
 Matches 129; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY	6689	accatcgatccaccacaaacgaaataaaagccaaatgcagggtttcatcagcgatggcga	6748
DB	1520	acaacgcgcatcgacaatgtggacggcaacgcgctgaggcatcgcccaagcgattgcaa	1579
QY	6749	tggcgctccatgcccaagcctacattcctcggcagatccatgggtaccgggggtatgccca	6808
DB	1580	ccgcaaggctcgttcaggcgatttgcgggcaagagatgatggcgatcgcgcgcgca	1639
QY	6809	cccaacacggtcaaggtgcggtggcagtggtgactgtcgaagctgtcggataatgggtcaat	6868
DB	1640	cttactcggcggaagcgggttaacgccatcggtactccagatttccgacggcggaatt	1699
QY	6869	gggtatttaaaatcaatggttcagcgcgatacccaagggccatgtaggggcgcgagttggtg	6928
DB	1700	ggattatcaaaaggcagcgttcggcaattcggcgccatttcggtgtcttcgcgcatctg	1759
QY	6929	caggttttcaact	6940
DB	1760	tcggttatcagt	1771

Search completed: September 12, 2001, 13:00:35  
 Job time: 1257 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2001, 03:34:48 ; Search time 18125.9 Seconds  
(without alignments)  
3620.328 Million cell updates/sec

Title: US-09-361-619-10  
Perfect score: 6942  
Sequence: 1 atgaatacacatctataaagt.....ttggtgcagggttttcacattt 6942

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est33:\*  
26: gb\_est34:\*  
27: gb\_est35:\*  
28: gb\_est36:\*  
29: gb\_est37:\*  
30: gb\_est38:\*  
31: gb\_est39:\*  
32: gb\_est40:\*  
33: em\_estba:\*  
34: em\_estfun:\*  
35: em\_esthum1:\*  
36: em\_esthum2:\*  
37: em\_esthum3:\*  
38: em\_esthum4:\*  
39: em\_esthum5:\*  
40: em\_esthum6:\*  
41: em\_esthum7:\*  
42: em\_esthum8:\*  
43: em\_esthum9:\*  
44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est41:\*  
111: gb\_est42:\*  
112: gb\_est43:\*  
113: gb\_est44:\*  
114: gb\_est45:\*  
115: gb\_est46:\*  
116: gb\_est47:\*

Thu Sep 13 14:18:05 2001

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
C 1	64.4	0.9	641	236	AQ946120	AQ946120 Sheared D
C 2	61.8	0.9	806	221	CNS04AEE	AL281759 Tetraodon
C 3	59.8	0.9	922	219	CNS0073K	AL066784 Drosophill
C 4	58	0.8	827	220	CNS02156	AL176451 Tetraodon
C 5	55.6	0.8	1101	219	CNS00LT2	AL078714 Drosophill
C 6	54.8	0.8	700	235	AQ940248	AQ940248 Sheared D
C 7	54.2	0.8	895	219	CNS0071A	AL066286 Drosophill
C 8	53.6	0.8	884	245	AZ550599	AZ550599 ENTCK07TR
C 9	53.4	0.8	880	107	AU088119	AU088119 AU088119
C 10	53.4	0.8	884	219	CNS006U0	AL065923 Drosophill
C 11	53.2	0.8	895	245	AZ546009	AZ546009 ENTFW53TF
C 12	53.2	0.8	905	245	AZ550256	AZ550256 ENTEV58TR
C 13	52.6	0.8	86	537	AU039635	AU039635 AU039635
C 14	52.6	0.8	540	107	AU034415	AU034415 AU034415
C 15	51.8	0.7	861	219	CNS0075A	AL066834 Drosophill
C 16	51.4	0.7	1163	220	CNS02GKK	AL196445 Tetraodon
C 17	51	0.7	541	221	CNS04AKU	AL195309 Tetraodon
C 18	51	0.7	681	220	CNS02EOD	AL193990 Tetraodon
C 19	50.8	0.7	568	219	CNS010PL	AL099267 Drosophill
C 20	50.6	0.7	1204	219	CNS016E2	AL106628 Drosophill
C 21	50.2	0.7	1101	219	CNS00PK2	AL077673 Drosophill
C 22	50	0.7	890	245	AZ530768	AZ530768 ENTBH54TF
C 23	49.8	0.7	853	220	CNS023KH	AL179594 Tetraodon
C 24	49.8	0.7	860	219	CNS018FL	AL109275 Drosophill
C 25	49.8	0.7	1014	222	CNS06CAJ	AL392273 T7 end of
C 26	49.6	0.7	1101	219	CNS00LO0	AL068607 Drosophill
C 27	48.8	0.7	415	23	A1645402	A1645402 v163801.y
C 28	48.2	0.7	447	107	AU088172	AU088172 AU088172
C 29	48.2	0.7	651	221	CNS03PNQ	AL254879 Tetraodon
C 30	48	0.7	575	218	AA550389	AA550389 1537m3 gm
C 31	48	0.7	621	236	AQ941683	AQ941683 Sheared D
C 32	48	0.7	1028	150	BF526646	BF526646 602070773
C 33	48	0.7	1101	219	CNS0100X	AL098379 Drosophill
C 34	47.8	0.7	452	137	BE579600	BE579600 kq31c06.y
C 35	47.8	0.7	937	219	CNS006ST	AL065880 Drosophill
C 36	47.2	0.7	791	219	CNS009KS	AL053801 Drosophill
C 37	47.2	0.7	900	245	AZ549980	AZ549980 ENTDD94TF
C 38	47.2	0.7	902	219	CNS006QP	AL065804 Drosophill
C 39	46.8	0.7	1101	219	CNS00PFX	AL071370 Drosophill
C 40	46.6	0.7	546	156	C22974	C22974 C22974 Dict
C 41	46.6	0.7	729	236	AQ945618	AQ945618 Sheared D
C 42	46.4	0.7	939	219	CNS00CNG	AL059400 Drosophill
C 43	46.4	0.7	975	219	CNS00T9I	AL074701 Drosophill
C 44	46.2	0.7	318	107	AU086474	AU086474 AU086474
C 45	46.2	0.7	500	107	AU087262	AU087262 AU087262

## ALIGNMENTS

RESULT	1	
AQ946120/c		
LOCUS		
DEFINITION		
	AQ946120	641 bp DNA
	Sheared DNA-46J23, TR sheared DNA	Trypanosoma brucei genomic clone
	Sheared DNA-46J23, DNA sequence.	
ACCESSION	AQ946120	
VERSION	AQ946120.1	GI:6769385
KEYWORDS	GSS.	
SOURCE	Trypanosoma brucei.	
ORGANISM	Trypanosoma brucei	
	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;	
	Trypanosoma	
REFERENCE	1 (bases 1 to 641)	
AUTHORS	El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M.	
TITLE	Determination of clone end sequences from Trypanosoma brucei GUTAT	

**JOURNAL  
COMMENT**

10.1 sheared DNA library  
Unpublished (1999)  
Other\_GSSs: Sheared DNA-46J23.TF  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: [neilsayed@tigr.org](mailto:neilsayed@tigr.org)  
Clones are derived from the Trypanosoma brucei GU7  
DNA library constructed at TIGR. Clones will be available  
distribution through ATCC. Sheared DNA end sequence  
<http://www.tigr.org/tdb/mbd/tbdb/>.  
Seq primer: M13-Reverse  
Class: shotgun

## FEATURES

[illegible]

[illegible]



```
Best Local Similarity 21.0%; Pred. No. 0.00013;
Matches 86; Conservative 140; Mismatches 184; Indels 0; Gaps 0;

QY 2507 ccaccactgaccacacaaatgctaagttaagcaaccaacttttagtaccaccgata 2566
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 905 mncmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcm 846

QY 2567 acgatcccttgttaacgcgaagacatcgccgaataatcaaccocctagcgaagaaa 2626
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 845 mncmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcm 786

QY 2627 ttacaccacaaagcacagcagacacgcctacacaaactttaagtaaaaagacg 2686
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 785 mncmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcm 726

QY 2687 gtgcaactgatgcgaacatccagctgggttaagatggtacacaaaacggaagacg 2746
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 725 mnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnm 666

QY 2747 tcaactctaaactcaaggtgaaacgggtctacgggtgctacaaataaagatggt 2806
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 665 amnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnm 606

QY 2807 cgttacccttggcattaacacacaaagcgtcttaagccggcgacagcaccactaa 2866
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 605 mncmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcm 546

QY 2867 acaaaatggtctgtattataaaaccccgctagtacgacaaatccaa 2916
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 545 raamcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcmcm 496

RESULT 4
CNS02156 827 bp DNA GSS 12-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
DEFINITION 224F10 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL176451.1 GI:7814508
VERSION GSS: genome survey sequence.
KEYWORDS Tetraodon nigroviridis.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
REFERENCE 1 (bases 1 to 827)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 827)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 827)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Direct Submission
JOURNAL Direct Submission
AUTHORS Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
TITLE This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
COMMENT http://www.genoscope.cns.fr/Tetraodon.
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source Location/Qualifiers
1..827
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
```

```
/clone="224F10"
/clone_lib="G"
/note="Genoscope sequence ID : COAG224DC05SP1-end :
PUC-Ori"
BASE COUNT 368 a 203 c 106 g 121 t 29 others
ORIGIN

Query Match 0.8%; Score 58; DB 220; Length 827;
Best Local Similarity 42.0%; Pred. No. 0.00037;
Matches 209; Conservative 19; Mismatches 270; Indels 0; Gaps 0;

QY 2213 gctatgcacacagacagggcaataatcaccagcagcgaagacaaatccaaagccg 2272
    ||||| : : : : : : : : : : : : : : : : : : : : : :
Db 126 GCTAACACAGCAGATTTAATGTGATCAGSAACAGCAGCAGCAGCAGCAGCAGCAG 185

QY 2273 ccagtatcggtgatattataataacagcgtttaaaccctaaataataagaactccgtg 2332
    ||||| : : : : : : : : : : : : : : : : : : : : : :
Db 186 ACAACAACAATACAAATACAGCAATACACAATAACACAACAACAACAACAACAACA 245

QY 2333 gcttgtctccacttaaacacactgttgactttatcgatggcaatgccaccacgcctaag 2392
    ||||| : : : : : : : : : : : : : : : : : : : : : :
Db 246 ACAATACACGRATACACAACAATAACACAACAACAACAACAACAACAACAACAACA 305

QY 2393 taacttacgatgaacaaatcaaacccagtaagtaacttatgatgtcaatgtgatgaga 2452
    ||||| : : : : : : : : : : : : : : : : : : : : : :
Db 306 ATACACAACAATAACAGCRATAACAACAATAACACAACAACAACAACAACAACAATA 365

QY 2453 aaaccattgaactcacagcgataatggcaagacaacaaatggcgtcaaaaccacca 2512
    ||||| : : : : : : : : : : : : : : : : : : : : : :
Db 366 ACAGCSATAACAACAATAACAGCVATAACAACAATAACAACAACAACAACAACAACA 425

QY 2513 cactgaccacacaaatgctaagtgttaagcaacaaacttttagtaccaccgataacgatg 2572
    ||||| : : : : : : : : : : : : : : : : : : : : : :
Db 426 ATACACAACAATAACACAACAATAACAGCGATACACATACAGCAGCGATGRCACRACR 485

QY 2573 ccttgttaacgcaagacatcgccgaataatcaaacccctagcgaagaaattcacca 2632
    ||||| : : : : : : : : : : : : : : : : : : : : : :
Db 486 ACAACAACAACAACAACAATAACAGCGATACAGCGRATAACAACAACAACAACAATA 545

QY 2633 ccacaaaggcacagcagacacgcgcctacaaaccttaagtcacaaagacggtgcaa 2692
    ||||| : : : : : : : : : : : : : : : : : : : : : :
Db 546 ACAATACACGGATGACACAACAATAACACAACAATAACAGTGATACACAACAATGACAGT 605

QY 2693 ctgatgcgaacacatca 2710
    : : : : : : : : : : : : : : : : : : : : : :
Db 606 ACACAAATGACACAACA 623

RESULT 5
CNS00LT2/c 1101 bp DNA GSS 14-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL078714
VERSION AL078714.1 GI:5102004
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Direct Submission
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
```

please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RC1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

FEATURES
  source
    Location/Qualifiers
      1. 1101
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone_11b="pCi-98"
        /clone="BACR48P19"
        /note="end : TEx3"
BASE COUNT      469 a      6 c      69 g      151 t      406 others
ORIGIN

```

Query Match 0.8%; Score 55.6; DB 219; Length 1101;  
Best Local Similarity 21.1%; Pred. No. 0.0018;  
Matches 115; Conservative 191; Mismatches 232; Indels 6; Gaps 2;

[illegible]

RESULT	6
AQ940248	
LOCUS	700 bp DNA
DEFINITION	Sheared DNA-42E21.TF Sheared DNA Trypanosoma brucei genomic clone
ACCESSION	AQ940248
	27-JAN-2000

VERSION A0940248.1 GI:6763513  
 KEYWORDS GSS.  
 SOURCE Trypanosoma brucei.  
 ORGANISM Trypanosoma brucei  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.  
 TITLE 1 (bases 1 to 700)  
 REFERENCE El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,  
 AUTHORS Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,  
 Fraser, C., and Adams, M.  
 DETERMINATION of clone and sequences from Trypanosoma brucei GUTat  
 10.1 sheared DNA library  
 Unpublished (1999)  
 OTHER\_GSSs: Sheared DNA-42E21.TR  
 CONTACT: Najib M. El-Sayed  
 DEPARTMENT of Eukaryotic Genomics  
 THE Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 TEL: 301 838 0200  
 FAX: 301 838 0208  
 EMAIL: nelsayed@tigr.org  
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared  
 DNA library constructed at TIGR. Clones will be available for  
 distribution through ATCC. Sheared DNA end sequences search page:  
<http://www.tigr.org/tdb/mdb/tbdb/>.  
 Seq primer: M13-Forward  
 Class: shotgun.

FEATURES	Location/Qualifiers	BASE COUNT	ORIGIN
source	1..700	374 a	
	/organism="Trypanosoma brucei"		
	/strain="TREU927/4 GUTat 10.1"		
	/db_xref="taxon:5691"		
	/clone="Sheared DNA-42E21"		
	/clone_lib="Sheared DNA"		
	/note="Vector: pUC18; Site_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."	43 c 41 g 242 t	

[illegible]



Matches 195; Conservative 0; Mismatches 214; Indels 3; Gaps 1;

1446 aattggttttagcaataaaagctggtagcagtgtgatgaacaaaaccttatcttgcataaga 1505  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
289 AGTAGATGATAGAAATTTTTGAAGAAAAGAAAAAAGAAAAANTCAGCAGAGCTTGAATTAGA 348  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1506 caagctaaaggttggcaacagcacccctaacaacggttgctgactgttaataaacaccat 1565  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
349 AATAAAAAAGATGACATCTTGTAATTACAAGCAATGGTATCATNAATCTTAACCAAAATCT 408  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1566 tgggtggtagcaataaacaatatcaagtcggtgctgatggcattaaatttg---ccgatgt 1622  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
409 TGCTGGATTACACACAGAAATTAATAAATAAAGAACAGAAATCATACTGAAACAGATAA 468  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1623 gaatgttaatgtaccaaatcgccaataatcgccactactcgtatttacocgaagaggaatat 1682  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
469 ATCTAATAAAGCATTACAGCTCAAAAGAGTTTATCAAGAACAAAAAGAAAAAATTAGANAAG 528  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1693 tggcttgctatctgatgttaaagttgatataaaagtcaccatatatttggataaaaaaaca 1742  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
529 TCAATGTGCTGATGTTTAATTAATTAANTAGACACAACTCAACAAGAACTGGTTGCTACTCA 588  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1743 acttcaaagtgggtggtgtlaaaattaccaaagacagcggcattaatgcaggtgcatcaaaa 1802  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
589 AGCTCGAGCTGATGGAATGAAAAGAAATTTGAGGATATTACTCAGGAACACAAATGGATG 648  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1803 gatcagtaattaaagatcaacggcagcgtaccgctcagtcagtcactcataaa 1854  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
649 GATTAGACAGCTTAAGANGCAGACTTAACAANTAGATTTCATTAAATFACTGAA 700  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9 EST 27-JAN-2001  
AU088119 500 bp mRNA AU088119 Sugano Malaria cDNA library Plasmodium falciparum cDNA  
OCCLUS clone XPFn6549, mRNA sequence.  
DEFINITION AU088119  
CCESION AU088119 GI:12390260  
KEYWORDS  
SOURCE malaria parasite P. falciparum.  
ORGANISM Plasmodium falciparum  
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
AUTHORS Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.  
TITLE FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum  
JOURNAL Nucleic Acids Res. 29 (1), 70-71 (2001)  
MEDLINE 20574754  
COMMENT Contact: Junichi Watanabe  
Institute of Medical Science  
The University of Tokyo, Department of Parasitology  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Tel: 81-3-5449-5378  
Fax: 81-3-5449-5410  
Email: jwatanab@nagame.ims.u-tokyo.ac.jp  
Suzuki,Y., Yoshimoto-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
.S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).

SOURCE 1. .500 Location/Qualifiers  
/organism="Plasmodium falciparum"  
/strain="3D7"  
/db\_xref="taxon:5833"  
/clone="XPFn6549"  
/clone\_lib="Sugano Malaria cDNA library"  
/dev\_stage="erythrocytic stage"  
BASE COUNT 214 a 37 c 86 g 163 t  
ORIGIN

Query Match 0.8%; Score 53.4; DB 107; Length 500;  
Best local Similarity 46.0%; Pred. No. 0.0052;  
Mismatches 211; Indels 0; Gaps 0;









---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.  
OM protein - protein search, using sw model  
Run on: September 13, 2001, 12:37:48 ; Search time 78.73 Seconds  
(without alignments)  
1781.835 Million cell updates/sec

Title: US-09-361-619-11  
Perfect score: 11694  
Sequence: 1 MNHKKVFNKATGTGTFMAVA.....NGSADTQGHVGAAGVGFHF 2314

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 50623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_0601.\*  
1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	11694	100.0	2314	22 AAB69136	M. catarrhalis les
2	7219.5	61.7	2053	22 AAB69135	M. catarrhalis str
3	6593.5	56.4	2047	22 AAB69134	M. catarrhalis str
4	6407.5	54.8	1992	22 AAB69133	M. catarrhalis str
5	6404.5	54.8	1992	17 AAW04505	Moraxella 200 kDa
6	6404.5	54.8	1992	22 AAB69137	M. catarrhalis M56
7	1177.5	10.1	2353	17 AAR99393	Haemophilus adhesi
8	1175.5	10.1	2411	21 AAB23860	Haemophilus influe
9	968	8.3	1104	21 AAB23856	Haemophilus influe
10	968	8.3	1104	21 AAB23859	Haemophilus influe
11	959.5	8.2	1004	21 AAB23857	Haemophilus influe

12	895.5	7.7	1002	21 AAB23854	Haemophilus influe
13	840.5	7.2	2039	19 AAW56322	Haemophilus paraga
14	796	6.8	2042	19 AAW56319	Haemophilus paraga
15	708	6.1	1094	21 AAB23858	Haemophilus influe
16	659.5	5.6	1098	17 AAR99392	Haemophilus adhesi
17	592	5.1	679	17 AAR99394	Haemophilus adhesi
18	592	5.1	679	21 AAB23855	Haemophilus influe
19	564.5	4.8	1601	18 AAW30292	Non-typeable Haemo
20	546	4.7	2514	21 AAV75097	Neisseria meningit
21	521	4.5	1598	18 AAW30291	Non-typeable Haemo
22	518	4.4	1529	14 AAR41732	High molecular wei
23	504.5	4.3	2599	21 AAY75098	Neisseria meningit
24	485	4.1	1536	18 AAW30293	Non-typeable Haemo
25	482	4.1	1536	14 AAR41723	High molecular wei
26	481	4.1	1536	14 AAR41725	High molecular wei
27	480	4.1	1536	15 AAR63505	Haemophilus high m
28	480	4.1	1536	21 AAB01846	Haemophilus influe
29	475	4.1	1477	14 AAR41724	High molecular wei
30	463.5	4.0	2383	21 AAB15945	E. coli proliferat
31	463	4.0	1477	15 AAR63506	Haemophilus high m
32	463	4.0	1477	18 AAW30294	Non-typeable Haemo
33	461	3.9	1477	14 AAR41728	High molecular wei
34	461	3.9	1477	21 AAB01848	Haemophilus influe
35	459.5	3.9	2893	19 AAW98828	H. pylori GHPO 148
36	459.5	3.9	2893	19 AAW71556	Helicobacter polyp
37	458.5	3.9	1638	20 AAY00138	Enterococcus faeca
38	458.5	3.9	1638	20 AAY00140	Enterococcus faeca
39	458.5	3.9	1638	20 AAY00142	Enterococcus faeca
40	452.5	3.9	1338	14 AAR41731	High molecular wei
41	448	3.8	3596	21 AAY87407	Bordetella pertuss
42	444	3.8	3647	11 AAR05041	Filamentous haemag
43	434	3.7	1981	19 AAW42634	Protein sequence t
44	433.5	3.7	2902	22 AAB46351	H. pylori HPN165 p
45	432	3.7	1188	21 AAB01844	Haemophilus influe

ALIGNMENTS

RESULT 1  
AAB69136  
ID AAB69136 standard; Protein; 2314 AA.  
XX AAB69136;  
XX AC  
XX 24-APR-2001 (first entry)  
XX  
XX M. catarrhalis les1 200kDa protein SEQ ID NO:11.  
DE  
KW Moraxella catarrhalis strain Q8; major outer membrane protein;  
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection.  
XX  
OS Moraxella catarrhalis.  
XX  
PN WC200107619-Al.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-CA00870.  
XX  
PR 27-JUL-1999; 99US-0361619.  
XX  
XX (CONN-) CONNAUGHT LAB LTD.  
XX  
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;  
XX WPI; 2001-159722/16.  
XX N-PSDB; AAF59105.  
XX  
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
XX useful in protective vaccines and for diagnosis -

PS Claim 1; Fig 5A-Y; 247pp; English.

XX The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. CC The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II), and its truncated versions, are used as CC immunogenic compositions and vaccines to protect against M. catarrhalis CC infections, particularly otitis media in humans. (II) is also used as CC antigen in immunoassays for detecting specific antibodies (Ab), and to CC generate Ab. (I) are used for recombinant production of (II) and its CC fragments are used as probes for identifying/cloning 200 kDa protein. CC genes from other strains, and for diagnostic detection of M. catarrhalis. CC (I) makes possible production of large amount of recombinant immunogens. CC Expression of truncated versions of (II) reduces toxicity of the protein CC towards the Escherichia coli host. The present sequence represents the CC M. catarrhalis lesi 200kDa protein, which is given in the exemplification CC of the present invention.

XX. Sequence 2314 AA:

Query Match 100.0%; Score 11694; DB 22; Length 2314;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHYIKVIFNKATGTFMAVEACAKSHSGSSSTAGQVGSSPVIRLTVATLAILVIGAT 60  
Db 1 mnhikvifnkatgtfmaveacakshsgssstaggvsspvirltrvatlailvigt 60  
Qy 61 LNSGAYAQNNKIAFGTTGNNDNASASNEASIAISLAKAHANOALAIIGSKPDPNRQAA 120  
Db 61 lngsayaqnnskiafgttgnndnasasneasiaislakahanoalaiigskpdpnrqea 120  
Qy 121 NOKAGSHAKGKESIAIGDVLAEGDASIAIGSDDLYLDRNSTNSKYPNGLLSTLIONHTV 180  
Db 121 nkeagshakgesiaigdvlaegdasiaigsdldyldrnstnskypngllstlioghtv 180  
Qy 181 LRQIRDSNGSQRYRTAEAGHASTAVGAMAYAKGHAFANAFGTRSTAEAGNYSIAVGLTAKA 240  
Db 181 lrqirdsngsqryrtaeaghasavgamayakghafanafgtrstaeagnysilevgitaka 240  
Qy 241 EKGVTIAIGSNAQAIIYNGALAGADTRVDLDYIGALGYGSGIILNNNNNNKAYVPEGNG 300  
Db 241 ekgytiaigsnaqaaiyngalagadtrvldldyigalgygsgilnnnnnnnkayvpegn 300  
Qy 301 SNIKSKATGNGLFSTGSGSTIKRKTIIVGAGYEDTDVAVNAQLKAVALAKRQITFKGDD 360  
Db 301 snikskatgnglfstgsgstikrktiivgagyedtdavvnaqlkavalakrqitfkqdd 360  
Qy 361 NGTGVKKLIGETLITKGGTQADKLTIDNNNIGVVDDNTGTLKVLAKNLSGLETVSTKNL 420  
Db 361 ngtgvkkligetlitkgggtqadkltidnnnigvvddntgtlkvlaknlsgletvstcknl 420  
Qy 421 TASEKTVVSGNNTAELOSGLTFPTTNAWSTDKTVYGTGGLKFTDNSNTALEDTTRITK 480  
Db 421 tasektvvsgnntaelosgltfptttnawstdktvvygtgglkftdnsntaledttritk 480  
Qy 481 DKICFSNKACTVDENKPYLDKDKLKYGNSTLNNNGGLTVNNTGGSNKQIQVAGDGIKAD 540  
Db 481 dkicfsnkatvdenkpyldkdklkygnstlnnnggltvnntggsknqiqvagdgiikad 540  
Qy 541 VNVNVSNAKFGTTRITEEIGFADADGKVDKSPYLDKKQLQVGVKTIKDSGINAGDQ 600  
Db 541 vnvvnvsnaakfgttriteeigfadadgkvdkspylldkkqlqvgvkkitkdsginagdq 600  
Qy 601 KISNVKATDDTDAVYKQLKQVQDADGALQSFSIRDEKGGQEFFISNYSNGTNPNTFE 660  
Db 601 kisnvkatddtdavykqlkqvqqdadgalqsfsirdekggqeffisnysngtntpntfe 660  
Qy 661 TITFAGENGISINDIAKGVKVICDIPINGLTTPKLTIVGSDKDKCTQLVIEQVAGSNDTRK 720  
Db 661 titfagengisindiakgvkvicdipinglttpkltivgskdkctqlvleqvagsndtk 720

Db 1801 nphagksgtdavalmniskavfkskdgtttttvssdgiisqgkdnssilcskdglnv9g 1860  
QY 1861 KVISNVGKGTDAANVOQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSNR 1920  
Db 1861 kvisnvkgkcktdaanvqqlnevrnllglnagnadnagnqvnadiakdpnsgssnr 1920  
QY 1921 TVIKAGTVLGGKNNDEKATLGGVGVGVDKDGANGDLSNVVVKTKDGSKKALLATYN 1980  
Db 1921 tvikagtvlggkgnndteklatgvgvgvdkdgnangdlsnvvvktgdkgskallatyn 1980  
QY 1981 AAGTNTLTNPAFAIDRINEOGIRFFPHVNDGNOEPVVOGRNGIDSSASGKHSVAIGFOA 2040  
Db 1981 aagtcnltmpaeaidrineogirffhvnndgnoepvvogrnidssasgkhsvaigfqa 2040  
QY 2041 KADGEAAVAIGROTQAGNQSIAIGDQAATCDOSIAIGTGNVVTGKHSGAIGDPSTVKAD 2100  
Db 2041 kadgeaavaigrtqagngqsaigldnaqatgqsaigtgnvvtgkhsaigdpstvkad 2100  
QY 2101 NSYSVGNNOFIDATQTDVFGVGNNTIVTESNSVALGSNAISAGTHIAGTQAKSGDTAG 2160  
Db 2101 nsysvgnnfdiatqtdvfgvgnntivtesnsvalgsnaisagthiagtgakskdgtag 2160  
QY 2161 TTTTAGATGVKGFAGOTAYGAVSVGASGAERRTIONVAAGEVSATSDAYVNGSOLYKATQ 2220  
Db 2161 ttttagatgtvkgfagotaygavsvgasgaerrtionvaagevsatscdavngsqlykatq 2220  
QY 2221 GIANATNELDHRIHONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGQGAVALG 2280  
Db 2221 gianatneldhrihonenkanagissamamasmpqayipgrsmvtggiathngggavag 2280  
QY 2281 LSKLSUNGQWFKINGSADTQGHVGAAGVGFHF 2314  
Db 2281 lsklsungqvwfkingsadtqghvgaavvgfhhf 2314

RESULT 2  
AAB69135  
ID AAB69135 standard; Protein; 2053 AA.  
XX AC AAB69135;  
XX DT 24-APR-2001 (first entry)  
XX DE M. catarrhalis strain Q8 200kDa protein SEQ ID NO:9.  
KW Moraxella catarrhalis strain Q8; major outer membrane protein;  
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection.  
XX OS Moraxella catarrhalis.  
XX PN WO200107619-A1.  
XX PD 01-FEB-2001.  
XX PF 26-JUL-2000; 2000WO-CA00870.  
XX PR 27-JUL-1999; 99US-0361619.  
XX PA (CONN-) CONNAUGHT LAB LTD.  
XX PI Loosmore SM, Sasaki K, Yang Y, Klein MH;  
XX WPI; 2001-159722/16.  
XX DR N-PSDB; AAF59104.  
XX PS New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
XX useful in protective vaccines and for diagnosis  
XX Claim 1; Fig 4A-V; 247pp; English.  
XX The present invention describes an isolated and purified nucleic acid (1)

CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.  
CC The 200 kDa outer membrane protein (II) has antibacterial activity and  
CC can be used in vaccines. (II), and its truncated versions, are used as  
CC immunogenic compositions and vaccines to protect against M. catarrhalis  
CC infections, particularly otitis media in humans. (II) is also used as  
CC antigen in immunoassays for detecting specific antibodies (Ab), and to  
CC generate Ab. (I) are used for recombinant production of (II) and its  
CC fragments are used as probes for identifying/cloning 200 kDa protein  
CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
CC (I) makes possible production of large amount of recombinant immunogens.  
CC Expression of truncated versions of (II) reduces toxicity of the protein  
CC towards the Escherichia coli host. The present sequence represents the  
CC M. catarrhalis strain Q8 200kDa protein, which is given in the  
CC exemplification of the present invention.  
XX  
SQ Sequence 2053 AA;  
Query Match 61.7%; Score 7219.5; DB 22; Length 2053;  
Best Local Similarity 64.1%; Pred. No. 0;  
Matches 1553; Conservative 140; Mismatches 250; Indels 481; Gaps 38;  
QY 1 MNHIYKVFENKATCTPMAVAPCAKSHSGSSSTAGOVGSSPVIRLTRVATLAILVIGAT 60  
Db 1 mnhiykvifnkagctfmaavaeyakshsgsgscatgvgsvrtisfariaalavivgat 60  
QY 61 LINGSAYAQN-NSKTAFTCTTGNNDN-----ASASNEASTAIGSLAKAHANOIAIGGSKPD 114  
Db 61 lngsayaqnttkleigtntkinnlkgdlatgeaslatgslskagsgaiaigsvkpd 120  
QY 115 PRQAAQKAGSHAKRESIAIGDVLAEGDASTAIGSDDLYLDRN-STNSKYNGLLST 173  
Db 121 p-nngsngnvgshaknesiaigddvlaegdastaignsddlylpknldlknef-----hk 174  
QY 174 LIONHTVLRIOIRDS-NGSQYRRTAAEGHASTAVGANAYAKHAFANAFGRSTAEKNSYL 232  
Db 175 lihghellkklqstdgklyrrtraghastavgamsyaqghfsnfaigtataeaysl 234  
QY 233 AVGLTAKAEKGYTTAIGSNAQAQINYGALALGADTRVDLDYGTALGYGSOILNNNNNNNK 292  
Db 235 avglaaatkqssiavgsnakanaafaataigtgnvtvnlgrvalgfgsqilidndntdas 294  
QY 293 AYVPEGNSNIKSSKATGNG-----LFSIG-----SSTIKRKIINVAGYEDTDVAVNAQL 343  
Db 295 ayvplgk-tladgkatrgqdstdfisgnnnnnssirrkilnvgagsrtdavnavql 353  
QY 344 KAVENLAKROITFK-GDDNGTGKVKKLGTPLTKGGETQADKLTDDNNNGVWTDNNTGL 401  
Db 354 klveelanrkitfkdgddnnnsverglnctltik-gdaqcnalte-anigvvcn-gl 410  
QY 402 KVKLAKNLGLTETVSTKNTLASEKVTYV-GSGNNTAEQLSGGLTFTPTTNASTDKTVYGTD 460  
Db 411 kvklakeltgitsvs-----atnkitysnntnnnaelqsggltsptgtktdktyvsid 465  
QY 461 GLKETDNSNT-ALEDTRITKDKIGFSNKAGTVDENKPYLDKDKLVGNSTLNGGLTVN 519  
Db 466 glkftdnsnsiatkgttrickkigfagtdngvdeskpyldneklkvgnstlsgsltcn 525  
QY 520 NFIGSSNKQIQVGADGKFKADVNVNVSN-AAKFGTTRTTEEEIGFADAGDGKDKSPYLD 578  
Db 526 ntfg--nkqigvganglkfatvannvantsatvgtariteekigfagtdngvdeqayld 583  
QY 579 KKOLOVGVKITKDSGINAGDQKISNVKDATDPTDVAITYKOLK----- 621  
Db 584 kerikvrveittdsginaghnkhtgltngiantdvtavikgkdkaptlnagdgisnnsn 643  
QY 622 ----- 623  
Db 644 ngdlvdsgnnttptynlsvkttklnsngtsgnnkfsvsnahdnnsivtakdladylnkv 703  
QY 624 QODADGALQSFISIR--DEKGOEFTISNLYSNGNTPNFTETITFAGENGISINDIAGKV 681  
Db 704 netadsalpsfkvgngdnnsnaltvgk-dtng---kftntklkgngvntnratgtv 759

Thu Sep 13 14:18:07 2001

QY	682	KVGDIPINGLTPKLTVGSDRGKTLQVIEQV--ASGNDTKNIIRGLSPITLPSITNAGGV	739
Db	760	tfigdsgnlttptkltvgdng-nrlviedqpsadgnstknlikglsptlpslaspsg-	817
QY	740	RTEGCGNTTTSDESKSAASIGDILITGFKLNKNSNSGVFVSTYNTVDFIDGNATTAKVT	799
Db	818	rniaignti-eekdknsaasldvlnagfnknngkdkdfvstydytdvdfidgnattatvt	876
QY	800	YDETNGTSKVTVVNVNDEKTIETLTDGNGTKNKIGVKYTTTLTTTANAKATNFSTDDNAL	859
Db	877	ydeangtskvaydvnnvdektietltdngk-kqlgvktikltetstngnattfstddhnl	935
QY	860	VNAKDIAENLTLAKELIHTTKTAOTAFLOTFKVKK-----DGATDDETTTVGKDGCTQNGKT	915
Db	936	vkasdiagnlntlaeeihctktantalcqtfvkvkvdendkaddtnaitykdkgt-sgk-	993
QY	916	VNTLKLKGENGLTVATNKDGTVTFTGINTOSGLKAGDSTTLNKGSLISKNPASNQIQVGA	975
Db	994	vntlklkgngldkdkdgtvtfgintqsgikagdstllnnglsikntasneqigvga	1053
QY	976	DGVKFAKVDKNSGSTGIDGTSRTIKDOIGFTGANGSLDTPKHLTKDKLKVGEVEITNTG	1035
Db	1054	dgvkfawvngvgagidgttritrdeigftgcnsgldkskphlskd-----g	1101
QY	1036	INAGGKITNTQSGDITQNSDNDVATGGRVYDLKTELESINSAAKTAQNSLHEPSVADEQ	1095
Db	1102	inaggkkitniqsgelaknshdvtgklydlktelenkissaktagnslhefsadeq	1161
QY	1096	CNHFTVSNPYSSTYDTSKTSVDITFAGENGITTKVNGVVRVGIDQTKGLTTPKLTVGNNN	1155
Db	1162	gnftvsnpyssydstsktsdvittfagengittkvngvrvvgidqtkgltpkltvgnnn	1221
QY	1156	GKIVIDSKGQNTITGLSNTLANVND-GAGHALSG-LANDTDKTRAASIGDVLNAGF	1213
Db	1222	kgivinsqngntitglsntlanvndkgsrvttedgnliikdedktraasivdvlsagf	1281
QY	1214	NLOGNEAVDFVSYDITVDIFDGNATTAKVTYDPTSKTSKVYDVNVNDRKTIETVSDKKL	1273
Db	1282	nlqngneavdfvstydytntfngantttakvtydstsktskvdyvndvdtiev-kdkkl	1340
QY	1274	GVKTTTLTKTSANGNATKFP---SAAQDGLVKASDIATHLNTAGDITQAKGASQASSA	1330
Db	1341	gvkttltstgfan--kfaisnqatgdalvkasdivahlnitsgdigtakgasqannsa	1398
QY	1331	SYVDADGNKVIYDSTDKKYQVNDKGOVDKNKEVAKDKLVAQAQTPDGTTLAQMNVKSVIN	1390
Db	1399	gyvdadgnkviydstonkyqvgakndgtvdktkevakkklvagaqtpdgtlaqmnvksvin	1458
QY	1391	KEQVNDANKKQGINEDNAFIKLENAAKDTKPKNAAVTVGDLNVAQTPLTFCAGDTGTTA	1450
Db	1459	keqvndankkqginednafvkglekaasdnktnaavtvgdlnavaqtpltfcagdtgltta	1518
QY	1451	KKIGETILTIKGGQTDINKLTNNIGVAGCTDGTFTVKLAKDLTNLNSVNAGTRIDEXGIS	1510
Db	1519	kkigetiltikgqtdcnkltndnigvagctdgtftvklakdltnlnsvnagtkidekgis	1578
QY	1511	FVDANGQAKANTPVLISANGLIDLGKRIISNIGAAVDDNDVANEKQFNEVAKVTNNLNQSN	1570
Db	1579	fvdangqakantpvlisangldl-----	1600
QY	1571	SGASLPPVTTDANGKPTNGTDPQKAIKADGKYHYHANANGVPVDKDGKPIITDADKLAN	1630
Db	1601	-----	1600
QY	1631	LAACHKPLDAGHQVVASIGGNSDAITLTNIKSTPLQIDTPNTGNAGQAQSLPSLSAAQ	1690
Db	1601	-----	1600
QY	1691	OSNAASVKDLNVGLENLQTNHINQVDFYKAYDTVNFVNGTGADITSVRSADGTMSTNTVNT	1750
Db	1601	-----	1600
QY	1751	ALAATDDGCVNLIIKADGKFYKADLMPNGSLKAGKSASDAKTPTGLSLVNPNAKGSTG	1810
Db	1601	-----	1600
QY	1811	DAVALNLSKAVFKSKDGTTTTTSVSSDGISIQKDNSSITLSKDLNVGCKVSNVKGKT	1870
Db	1601	-----gqkvlsnvvgkt	1612
QY	1871	KUTDAANVOQLNEVRNLLGLGNAGNDNADGNQVNTADIKKDPNSGSSSNRTVIKAGTVLG	1930
Db	1613	ktdaanvqqlnevrnllgl---gndnagdnqvnladiikkdpnsgsssnrtvikagtvlg	1669
QY	1931	GKGNNDTEKLATGGVQGVYDKDGNANGDLSNVWYTKDQGSKKALLATYNAAGOTNYLTN	1990
Db	1670	gkgnndteklatggvqgvdkdgnangdlnsvwtkdsgskallatynaagqcnvytn	1729
QY	1991	NPAAEDIRINEQIRFFHVNDGNQEPVQVGRNGIDSSASGKHSVAIGFQAKADGEEAAVAI	2050
Db	1730	npaaedirinedgirffhvndgnqepvqvgvrgngidssasgkhsvaigfqakadgeaava1	1789
QY	2051	GROTAQGNOSIAIGDNAQATGDOSTAIGTGNVVTGKHSGAIGDPPSTVKADNSYSVGNNO	2110
Db	1790	grtqagnqsiaigdnaqatgdqslaigtgnvvvagkhsaigdpstvkdnsysvgnnnq	1849
QY	2111	FIDATQDFVFGVGNNTVTFESNSVALGSNSATSAGTHACTQAKKSDGTAGTTTACATGT	2170
Db	1850	ftdatqtdvfvgvgnntvtfesnsvalgsnsatsagthactqakksdgtagtttttagatgt	1909
QY	2171	VKGAGOTAVGAVSVGASGAERRIQNVAAGEVSATSTDAVNGSOLYKATQGTANATNELD	2230
Db	1910	vkqagotavgavsvgsgaerriqnvaagevsatstdavngsqlykatqgsana cnelid	1969
QY	2231	HRHQENKANAGLSSANAMASMPQAVIPGRSMVTGGIATHNGQGAVALGSLKSLSDNGW	2290
Db	1970	hrihqenkanaglsanamasmpqayipgrsmvtggiaathngqgavavglsklsdngw	2029
QY	2291	VFKINGSADTQGHVGAAGVAGFHF	2314
Db	2030	vfkingsadtqghvgaavagfhhf	2053
RESULT 3			
AAB69134 standard; Protein; 2047 AA.			
XX	AAB69134;		
AC	AAB69134;		
XX	24-APR-2001 (first entry)		
DT	M. catarrhalis strain 4223 genomic 200kba protein SEQ ID NO:7.		
XX	Moraxella catarrhalis.		
XX	Moraxella catarrhalis.		
OS	WO200107619-A1.		
PN	01-FEB-2001.		
XX	26-JUL-2000; 2000WO-CA00870.		
PF	27-JUL-1999; 99US-0361619.		
XX	(CONN-) CONNAUGHT LAB LTD.		
XX	Loosmore SM, Sasaki K, Yang Y, Klein MH;		
PI	WPI; 2001-159722/16.		
XX	N-PSDB; AAF59102, AAF59103.		
DR	New nucleic acid encoding Moraxella catarrhalis outer membrane protein,		
PT			



Thu Sep 13 14:18:07 2001

QY 1699 DVLNVCFNLQNHNOVDFVKAYDTVNFVNGTGADITSVRSADGTMSTNTVNTALAAATDDDD 1758  
Db 1592 ----- 1591  
QY 1759 GNVLLKAKDKRFYKADLLMPNGLKAGKSASDAKTPGTGLSLVNPAGKSTGDAVALNNL 1818  
Db 1592 ----- 1591  
QY 1819 SKAVFKSKDGTITTTVSSDGISIQGKDNSSITLSKDLGNVGKGVISNVGKGTGKTDAAV 1878  
Db 1592 ----- 1591  
QY 1879 QQLNEVRNLLGLNAGNDNADGNQVNIADIKDPNSGSSNRTVTKAGTVLGGKGNNDTE 1938  
Db 1612 qqlnevrnllglnagnndnagnqvnidiakdpnsgssnrtvlikagtvlggkgnndte 1671  
QY 1939 KLATGCVQGVVDKGNANGDLSNVVVKQKDGSKKALLATYNAAGQTNLYLNNPAAEADR 1998  
Db 1672 klatgglqvvdkgdngangdlsnvvvktgdkgskallatynaagqtnlylnnpaealdr 1731  
QY 1999 INEQGTRFHVNDGNQEPVVOGRNGIDSSAGKHSVAIGFQAKADGEAAVAITGROTAGN 2058  
Db 1732 ineqgtrfhrvndgnqepvvogrnngidssasgkhsvaigfqakadgeaavaigrtqtqgn 1791  
QY 2059 QSTAIQDNQAQATGDSIAIGTGNVVTGKHSGAIGDPSVTVKADNSYVGNNGQFIDATOTD 2118  
Db 1792 qstaiqdnqaatgdsiaigtgnvvagkhsgaigdpstvkdnsyvgngnngftdatqtd 1851  
QY 2119 VFGVGNITVTESNVALGSNSAISACTHAGTQAKKSDGTAGTTPAGATGVKGFAGQT 2178  
Db 1852 vfgvgnitvtesnvalgsnsaisacthagtkakksdgtagtttagatgvlkvfgagqt 1911  
QY 2179 AVGAVSAGSAGERRIQNVAAGEVSATSDAVNGSOLYKATQCIANATNHELDRHIONEN 2238  
Db 1912 avgavsgsagerriqnvaaagevsatstdavngsolykatqcianatneldhrhionen 1971  
QY 2239 KANAGISSAMAMSPQAYIPGRSMVTGGTATHNGGCAVAVGLSKLSDNGQWVFKINGSA 2298  
Db 1972 kanagissamamspqayipgrsmvtggatathnggcavavglsklsgdngqvwfkingse 2031  
QY 2299 DTQGHVGAAGVAGGFHF 2314  
Db 2032 dtqghvgaagvghf 2047  
RESULT 4  
ID AAB69133 standard; Protein; 1992 AA.  
XX  
AC AAB69133;  
XX  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kda protein SEQ ID NO:3.  
XX  
XX Moraxella catarrhalis strain 4223; major outer membrane protein;  
KW 200kda outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection.  
XX  
OS Moraxella catarrhalis.  
XX  
PN W0200107619-A1.  
XX  
XX 01-FEB-2001.  
XX  
XX 26-JUL-2000; 2000WO-CA00870.  
XX  
XX 27-JUL-1999; 99US-0361619.  
XX  
PA (CONN-) CONNAUGHT LAB LTD..  
XX  
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;  
PI

XX WPI; 2001-159722/16.  
DR N-PSDB; AAF59100, AAF59101.  
XX  
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
PT useful in protective vaccines and for diagnosis  
XX  
XX Example 3; Fig 2A-W; 247pp; English.  
PS  
XX  
XX The present invention describes an isolated and purified nucleic acid (I)  
CC that encodes a 200 kda outer membrane protein of Moraxella catarrhalis.  
CC The 200 kda outer membrane protein (II) has antibacterial activity and  
CC can be used in vaccines. (II), and its truncated versions, are used as  
CC immunogenic compositions and vaccines to protect against M. catarrhalis  
CC infections, particularly otitis media in humans. (II) is also used as  
CC antigen in immunoassays for detecting specific antibodies (Ab), and to  
CC generate Ab. (I) are used for recombinant production of (II) and its  
CC fragments are used as probes for identifying/cloning 200 kda protein  
CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
CC (I) makes possible production of large amount of recombinant immunogens.  
CC Expression of truncated versions of (II) reduces toxicity of the protein  
CC towards the Escherichia coli host. The present sequence represents the  
CC M. catarrhalis strain 4223 lambdaEMBL3 clone 200kda protein, which is  
CC used in the exemplification of the present invention.  
XX  
XX Sequence 1992 AA:  
SQ  
Query Match 54.8%; Score 6407.5; DB 22; Length 1992;  
Best Local Similarity 60.5%; Pred. No. 4.2e-275;  
Matches 1428; Conservative 146; Mismatches 316; Indels 471; Gaps 48;  
QY 56 VIGATLNGSAYAQ--NNSKTAFTGTG---NNDNASASNEASIAIGSLAKAHAAQATAIGG 110  
Db 1 vigaatlsgsayaqkdkthiaigeqnqprsrsgtakadgdraiaigenanagaqggalaigs 60  
QY 111 SKPDPRNOAQAQAGSIAIGDVLAEADGASIAIGSDDLYLDRNSTSKYKPN-- 169  
Db 61 snktv-ngssldkigtatqesialgdkasdaslaigsdldhldghgnpkhpgt 119  
QY 170 LLSTLIQNHVTLRQIRDSNGSQ-KYRRTAEGHASTAVGAMAYAKHFANAFGRSTAE 228  
Db 120 lndlinghavlketrskdndvkryrrttasghastavgamsyegghfsnafgrctaks 179  
QY 229 NYSLAVGLTAKAEKYTIAIGSNAQAINYGALAGADTRVDLYGIALGYSQILNNNN 288  
Db 180 ayslavglataegstiaigsdatssalgaialgagraqlggsialgqgs-vvtqsdn 238  
QY 289 NNNKAYVPEGNSNISKSKATGN---GLFSGTSSTIKKILNVGAGYEDTAVNVAOLKA 345  
Db 239 nsrpaytpntqaldpk-fqatnntkagplsignsikrklinvgagvnxktadvnvaqla 297  
QY 346 VENLAK-ROITFKDDNGTGKVKLGETITIKGG-ETQADKLTDDNNIGVYTD-NNTGLK 402  
Db 298 vkwakerrittqgdnstcdvkgidntltikgaetna--ltd-nnigvkeadnsglk 354  
QY 403 VKLAKNLGLETYSTKNLTAASEKVTVGSGNN-TAELOSGLTFT-PTTNA-STDTVYGT 459  
Db 355 vkiaktlnlntevnttlnattvkvgsssttaellsdslftqntqsgstaktyv 414  
QY 460 DGLKFTDNS-NTALEDTRITTKDKIGFSNKAQTVDENK-PYLDKDKLKVGNSTLNGG 517  
Db 415 ngvktfnnaettaagtrtrtrtdkigfa-rdgvdvdeqapyldkklkvgsv----ait 468  
QY 518 VNNTIGGSNKQIOVGADGKIFAD-VNVVNSNAKFTTRITEEIGFADAGCKVKSPY 576  
Db 469 idngidagnkkslnlakgssandavtieqlkaak--ptlnagaglsvtptelsvdksgn 526  
QY 577 LDKKQLOVGGVKTIKDSGINAGDQKISNVKDATDDTDAVTYKQ----LKQVQODADGALQ 632  
Db 527 vtatyni-gvkttelnsdgttsdk--fsvkgsgttnnslvtaehlasylnevnrtadsalq 583  
QY 633 SFSIRDEKQGEFTISNLYSNGNTPN--TFETITFAGENGISINDIAKGVKVGIDP 690



```
Db 584 stvkeedaddanaicv-akdttknagavsiiklkgkngltvatk-kdgtvfglsqdsq 641
QY 691 LTPPKLTVGSD---KDGKTLQVIEQVAGSNDTKNT- 722
Db 642 ltigketlmdglvtkdneq-iquvangikftnvngsnpgtgiantariirdklfgag 699
QY 723 -----IRGLSPTLPSTNAGGVRTTEQGN 747
Db 700 sdgavdtnkpyldqklqvgnvkitntgtnaggaikgtlpsladqss-rnieignt 758
QY 748 IYSDDEKSAASIGDLTNGFNLKNNNSVGFVSTYNTVDFIDGNATTAKVIYDETNTQS 807
Db 759 i-qdkksnaasindilnlgfalknnonpldvstydvlfdangnattatvthdtanks 817
QY 808 KVTYDYNVDEKTELTG-DNGKTKNKIGVKTTLITTTNANGK-ATNF--SVTNDALVNAK 863
Db 818 kvvydvrvaddtthigtadnk--klgvtkltnksangntatnfnvnssdedalvna 875
QY 864 DIAENLTLAKEIHTTKGTADTALOTFKYK---DGATDDEITVVGDKGTQNGKTVNTL 919
Db 876 diaenlntlakeihtekgtadaltqftvkvdennaddanaicvvgknann--qvntl 933
QY 920 KLKGENGLVATNKDGTVTFGINTQSLKAGDSTTLNKDGLSKKNPASNEQIQVGADGVK 979
Db 934 tlkgenglniktdkngvtfgintsglkagks-tlndgglsiknptgseqiqvgadvk 992
QY 980 FAKV-DKGNSSGTCIDGTSRTIKDOIGFTGANGSLDPTKPHLTDKDLKVGVEITNTGINA 1038
Db 993 fakvnnngvvagidgttridtrdeligtungslskphlskd-----glna 1040
QY 1039 GKKITNIQSGDITQNSDVAVTGGRYVDLKTLESKINSAAKTAQNSHEFSVADPQGNH 1098
Db 1041 gkkitniqsgelagshdavgtkyldiktelenkisstaktagnslhefsvadeqgn 1100
QY 1099 FTVSNPYSYDTSKSDVITFAGENCITTKVKNKGVVRVIGDQTKGLTTPKLVGNNGKG 1158
Db 1101 ftsvnpysydsksdvitfagengitckvknvrvvgldqtkgtltpkltvgnnngkg 1160
QY 1159 IVIDSKDQNTITGLSNTLANVTND--GAGHALSQS--LANDTOKTRAASITGDVLNAGFNQ 1216
Db 1161 ividsqngqntlitlantlanvntndkgsvrtteqgnlikdedktraasivdvlseagfnlq 1220
QY 1217 GNGEAVDFVSTYDVFIDGNATTAKVTYDDTSKTSKVYDVNDVNDKTEIYTSKKLGVK 1276
Db 1221 gngeavdfvstydvfnfadgnattakvttydtsktskvvydvndvndttiev-kdkklgvk 1279
QY 1277 TTTLTTKTSANGATKE--SAADGDALVKASDIATHLNTLAGDIQTAKGASQASSASYV 1333
Db 1280 tttltstgtgan--kfalsnqatgdalvkasdlvahnltsgdqitakgasqannsgayv 1337
QY 1334 DADGNRTYDSTDKYYQVNDKGVDPKNEKAKDKLVAQAQTPDGTLAGMNVKSVINKEQ 1393
Db 1338 dadgnkviydstdnkyyqakndgtvdktevakdklivaqaqtpdgtlagmnavksvinkeq 1397
QY 1394 VNDANKQGINEDNAPIKGLENAKDKTKYNAAVTVGDNLNAVAQTPLTFAGDTGTTAKKL 1453
Db 1398 vndankkqginednafvgleaasdnktnaavtvgdlnavaqtpiltfagdtgttakk1 1457
QY 1454 GETLTITKGGOTDPTNKLTDDNIGVAVAGTDGFTVKLAKDLNLNSVNAGGTRIDEKGSFVD 1513
Db 1458 getltikggotdnkltddnigvavldgftvklakdlnlsvnaggtklddkgvsfvd 1517
QY 1514 ANGOAKANTPVLISANGLJDLGGKRISNIGAIVDDNDVAFNKQFNEVAKTVNNLNNQNSGA 1573
Db 1518 ssgakantpvlisangldl----- 1536
QY 1574 SUPFVVTDANGKPIINGTDGKPQKAIKAGDKYYHANANGVPVVDKDGKPIPTDADKLANLAA 1633
Db 1537 ----- 1536
QY 1634 HGKPLDAGHOVVASLGGNSDAITLTNIKSTLPQIDTPTNTGNANAGQAQSLPSLSAAQSN 1693
```

```
Db 1537 ----- 1536
QY 1694 AASVKDVLNVGNFLQTNHNOVDFVKAYDTVNFVNGTGADITSVRSADGTMNTVNTALA 1753
Db 1537 ----- 1536
QY 1754 ATDDGCVLIKAKDGKFKYRADDLMPNGSLKAGKSASDAKPTPTGLSLVNPWAGKSGTDAV 1813
Db 1537 ----- 1536
QY 1814 ALNNLSKAVPKSKDGTITTVSSDGLSIQCKDNSSITLSKDCLNVGCKVLSNVCKTKDT 1873
Db 1537 ----- 1551
QY 1874 DAANVQQLNEVRNLLGLGNAGNDNADQNVNIADIKKDPNSGSSSNRTVTKAGTVLGGKG 1933
Db 1552 deavnvqlnevrnllglgnagnadnagnvniadikkdpngsgssnrtvlikagtvlgkg 1611
QY 1934 NNDTEKLATGGVGVVDKDGANGDLSNVWVKTKDGSKKALLATYNAAGOTNYLTNPA 1993
Db 1612 ndteklatggigvgvdkdgnangdlisnvwvktqkgskkalllatynaagqtnyltnpa 1671
QY 1994 EADIRINEQIRFFHYNDGNOQEPVVOGRNGIDSSASGKHSVAIGFQAKADGEAAVALGRQ 2053
Db 1672 eadrineqgirffhvdngnqepvvqgrngidsasgkhsvaigfqakadgeaavaigrq 1731
QY 2054 TOAGNOSIAIGDNAQATGDSQSTAIGTGNVVTGKHSAGIAGDPSTVKADNSYSVGNNOQFID 2113
Db 1732 tqagnqsiaigdnagtgdqsgiaigtgnvvvaghkhsaigdpstvkadnsysvgnnqftd 1791
QY 2114 ATQTDVFGVGNNTVTESNSVALGNSAISAGTHAGTQAKKSDGTAGTTTAGATGTYVK 2173
Db 1792 atqtdvfgvgnntvtesnsvalgnsaisagthagtkaksdgtagtttagatgtvk 1851
QY 2174 FAGOTAVGAVSVCASCAERIQNVAAAGEVSATSTDAVNGSOLYKATQGIANATNELDHR 2233
Db 1852 fagotavgavsvgasgaerriqnvaagevsatstdavngsqlykatqgsianatneidhr1 1911
QY 2234 HQENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGQGAVAVGLSKLSNGQWVFK 2293
Db 1912 hqenkanagissamamampqayipgrsmvtggiathngqgavavglsklsdngqvwfk 1971
QY 2294 INGSADTQGHVGAAGAGPHF 2314
Db 1972 ingsadtqghvgaavgagfhf 1992
RESULT 5
AAW04505
ID AAW04505 standard; Protein; 1992 AA.
XX
AC AAW04505;
XX
DT 25-JAN-1997 (first entry)
XX
DE Moraxella 200 kDa outer membrane protein.
XX
KW Outer membrane protein; OMP; immunogen; vaccine; otitis media;
KW diagnosis.
XX
OS Mycobacterium catarrhalis strain 4223.
XX
PN W09634960-A1.
XX
PD 07-NOV-1996.
XX
PF 29-APR-1996; 96WO-CA00264.
XX
PR 26-MAR-1996; 96US-0621944.
PR 01-MAY-1995; 95US-0431718.
PR 07-JUN-1995; 95US-0478370.
XX
PA (CONN-) CONNAUGHT LAB LTD.
```



Db 1537 ----- 1536  
QY 1754 ATDDGNVLIIKAKDKFYKADLMPNGSLKAGKSASDAKPTTGLSLVNPNAKGSTGDAV 1813  
Db 1537 ----- 1536  
QY 1814 ALNNLSKAVFKSDGTTTTTSSDGSISIOCKDNSSITLSKDGILNVGKVTSNVKGTKDT 1873  
Db 1537 ----- 1551  
QY 1874 DAANVOOLNEVRLLGLGNAGNADNOVNIADIKDPNSGSSSNRTVTKACTVLGGKG 1933  
Db 1552 daanvqilnevrllglgnagnadnqvnliadikdpnsgsssnrtv.kagvtvlggk 1611  
QY 1934 NNDEKATGQGVGVGDKNANGDLSNVWVKQDKGSKKALLATYNAACOTVLTNNPA 1993  
Db 1612 ndteklatggiqgvdkdnangdlsnvvvtqkdgskkallatynaagotvyltnpa 1671  
QY 1994 EADRINEQIRFHVNDGNQEPVVOGRNGIDSASGKHVAIGFQAKADGEAAVATGRQ 2053  
Db 1672 eaidrineqgirffhvnqepvvggrngidsasgkhsvaigfqakadgeaavaiqrq 1731  
QY 2054 TOAGNOSIATGDNAQTGDSIAIGTGNVYTKHSGAIGDPSTVKADNSYSGVNNQFID 2113  
Db 1732 tqagngsiaigdnhaqatgdsiaigtgnvvagkhsaigdpstvkadnsysvgnnqftd 1791  
QY 2114 ATQDVFVGNNITVTFESNVALGSNSAISAGTHAGTQAKKSDGTAGTTTATAGTCVKG 2173  
Db 1792 atqtdvfvgnnitvtfesnvalgsnsaisagthagtqakksdgtagttttagatgtvk 1851  
QY 2174 FAGQAVGAVSGASGAERRIQNVAGEVSATSDVNGSOLYKATQGIANATNEIDHRI 2233  
Db 1852 fagqtavgavsgasgaerriqnvagevsatstdvngsolykatqgsianatneidhri 1911  
QY 2234 HONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGGQAVAVGLSKLSDNGQWVFK 2293  
Db 1912 hqnenkanagissamamaspqayipgrsmvtggiaathnggqavavglsklsdngqvwfk 1971  
QY 2294 INGSADTQGHVGAAGVAGPHF 2314  
Db 1972 ingtsadtqghvgaavgagphf 1992  
RESULT 6  
ID AAB69137  
XX AAB69137 standard; Protein; 1992 AA.  
AC AAB69137;  
XX 24-APR-2001 (first entry)  
XX M. catarrhalis M56 200kDa protein in pKS348 SEQ ID NO:13.  
XX Moraxella catarrhalis strain Q8; major outer membrane protein;  
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
KW otitis media; detection.  
XX Moraxella catarrhalis.  
OS WC200107619-A1.  
XX 01-FEB-2001.  
XX 26-JUL-2000; 2000WO-CA00870.  
XX 27-JUL-1999; 99US-0361619.  
XX (CONN-) CONNAUGHT LAB LTD.  
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;  
PI WPI; 2001-159722/16.  
XX DB

DR N-PSDB; AAF59106.  
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
PT useful in protective vaccines and for diagnosis  
XX Claim 1; Fig 8A-V; 247pp; English.  
PS The present invention describes an isolated and purified nucleic acid (I)  
XX that encodes a 200 kDa outer membrane protein (II) has antibacterial activity and  
CC The 200 kDa outer membrane protein (II) has antibacterial activity and  
CC can be used in vaccines. (II), and its truncated versions, are used as  
CC immunogenic compositions and vaccines to protect against M. catarrhalis  
CC infections, particularly otitis media in humans. (II) is also used as  
CC antigen in immunoassays for detecting specific antibodies (Ab), and to  
CC generate Ab. (I) are used for recombinant production of (II) and its  
CC fragments are used as probes for identifying/cloning 200 kDa protein  
CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
CC (I) makes possible production of large amount of recombinant immunogens.  
CC Expression of truncated versions of (II) reduces toxicity of the protein  
CC towards the Escherichia coli host. The present sequence represents the  
CC M. catarrhalis M56 200kDa protein in pKS348, which is given in the  
CC exemplification of the present invention.  
XX Sequence 1992 AA;  
SQ  
Query Match 54.8%; Score 6404.5; DB 22; Length 1992;  
Best Local Similarity 60.4%; Pred. No. 5.6e-275;  
Matches 1427; Conservative 147; Mismatches 316; Indels 471; Gaps 48;  
QY 56 VIGATLNGSIYAAQ--NNSKIAFGTTG---NNDNASASNEASIAIGSLAKAHANQAIAGG 110  
Db 1 migatlngssyaqkdkthiaigeqnqprsgtakadgdraiaigenanagggaiaigs 60  
QY 111 SKDPDRNQAANQKAGSHAKGESIAIGDVLAEQDSIAIGSDDLYLDRNSTNSKYPNG- 169  
Db 61 snktv--nsidklgtatgqesiaigddvkasdasiaigsdldhldqngnphkpgt 119  
QY 170 LLSTLIQNHVTLRQIRDSNGSQ--KYRTAAEGHASTAVGAMAYAKGHAFANAFGRSTAE 228  
Db 120 lindlinghavlkerrskdndvkyrrttasghastavgamsyaqghfsnagftrataks 179  
QY 229 NYSLAVGLTAKAEKGYTIAIGSNQAQINYGALGADTRVDLDYGIAGVGSGQLNNNNN 288  
Db 180 ayslavglataeagqstiaigsdatssslgaialagatraqlgssialgggs-vvtqsdn 238  
QY 289 NNNKAYVPEGNSNIKSSKATGN--GLFSTGSSITKRIINVGAGYEDTDVAVNAOLKA 345  
Db 239 nsrpaytqtdqldpk-fgatnntkagplsgnsikrkinvgagvntkdavnavqla 297  
QY 346 VENLAK-RQITFRKDDNGTGVKKKLGELTLTKGG-ETQADKLTDDNNNIGVVTD-NNTGLK 402  
Db 298 vkwakeritfgddndstdvkigldntltikggaetna--ltd-nnigvvkeadnsglk 354  
QY 403 VKIAKNSLGSETVSTKNLTASEKVTYVSGNN-TAELOSGLTFT-PTNA-STDKTYVGT 459  
Db 355 vkiaaktnnltvntttlnattvkvgssssttaellsdslftqptgsgstsktyvgv 414  
QY 460 DGLKFTDNS-NTALEDTTRITKDKIGFSNKGATVDENK-PYLDKDKLVGNSLTNNGLT 517  
Db 415 ngvkttnnaettaalgttrtdrkigfa-rdgdvdekqapylkdkkqkvgsv-----ait 468  
QY 518 VNNITIGGSNKQIOVGADGIKFAD-VNVNVSNAAKFGTTRITEEBEIGFADADGKVDKKS 576  
Db 469 idngidagnkkslnlagssandavtieqlkaak--ptlnagagisvtptseisvdaksgn 526  
QY 577 LDKKQLQVGVGKTKSGINAGDQKISNVKDATDDTDAVTYKQ----LKQVQDADGALQ 632  
Db 527 vtaptyni-gvkttelnsdgtssdk--fsvksgsgnnsilvtaehlasylnevrntadsalg 583  
QY 633 SFSIRDEKGOEFTISNLYSNGNTPN--TFETITPAGENGISINDIAKGVKVCIDIPID 690  
Db 584 sftvkeedddanaity-akdttnagavsilklkngitvatck-kdgtvtflglsqdsq 641

QY 691 LTPKLTGVSD-----KDGKTLQVIEQVASCNGTKNI----- 722  
Db 642 lligkstinndgltvkdtneq--lqvganglkftvnngsnpgtgiantaritrkdigfag 699  
QY 723 -----IRGLSPTLSITNAGGVRTTEOGNT 747  
Db 700 sdgavdtnkpyldqklqvgnvkntntginaggkaigtglstplpsiadqss--rnielgnt 758  
QY 748 ITSDEKSKAASIGDILNTGFLKNNSNSVGFVSYNTYVDFIDGNATTAKTYDETQTS 807  
Db 759 i-qdkaksnaasindilntgfnlknnpidfvstydivdfangnattatvthdLankts 817  
QY 808 KTYDVNVNDEKTIETG--DNGTKNIGVKTTLTTLTYNNANGK--ATNF--STTDNDALVNAK 863  
Db 818 kvvydvnvddtclhlgtddnk--klgvttklntsangntatfnvnnsdedalvnaK 875  
QY 864 DTAENLTLAKETHHTKGTADTALOTFKVK-----DGATDDDETITVVGKDGTONGKTVNTL 919  
Db 876 diaenlntlakehttkgtadaltqftvkvdennnaddanaicvgkknann--qvntl 933  
QY 920 KLKGENGLVATNKGDTVTFGINTOSGLKAGSDTTLNKDGLSIKPNASNEQIQVGADGVK 979  
Db 934 tlkgenglnlktkngtvtfginttsglkagks--tlndgglsiknptgseqiqvgadgvk 992  
QY 980 FAKV--DKGNSSTGIDGTSRTITDOIGFTGANGSLDTPPKPLTKDKLKVGEVEITNTGINA 1038  
Db 993 fakvnnngvvgagldgtrtrdelgftgungslskspklskd-----gina 1040  
QY 1039 GKKITNIQSGDITQNSDAVTCGRVYDLKTELESKINSAAKTAONSLHEFSVADEQGNH 1098  
Db 1041 gkkitniqsgaiaqshdvtgklydltelenkisstaktagnslhefsvadeqgn 1100  
QY 1099 FTVSNPVSSTDTSEKTSVITFACENGITTKVNGVVRVGIDOTKGLTTPKLTGVNNGNGK 1158  
Db 1101 fTVSNPVSSTDTSEKTSVITFACENGITTKVNGVVRVGIDOTKGLTTPKLTGVNNGNGK 1160  
QY 1159 IVIDSKGQNTITGLSNTLANVND--GACHALSGQ--LANDTDKTRAASIGDVLNAGFNLO 1216  
Db 1161 lvidsqngqntitglntlanvndkgsvrttegnlikdedktraasivdvlsagfnliq 1220  
QY 1217 GNGEAVDFVSTYDTPVDIDGNATTAKTYDDTSEKTSKVYDVNVNDKTEVTSDKKLGVK 1276  
Db 1221 gngeavdfvstydvntfnadgnattaktyddtsektkskvvydvnvddttiev--kdkklgvk 1279  
QY 1277 TTTLTKTSANGNATKF---SAAEDDALVKASDIATHLNTLAGDIOTAKGASQSSASVY 1333  
Db 1280 tttltstgtgan--kfalsngatgdalvksadlvahIntlsgdiqtakgasqansagvy 1337  
QY 1334 DADGNKVITYDSTDKKYQVNDKGOVDKNKEVAKDKLVAQAOTPDGTLAOMNVKSVINKEQ 1393  
Db 1338 dadgnkvlydstdknyyakndgtvdktkvakdklvagaqtpdgtlaqmnvksvinkeq 1397  
QY 1394 VNDANKKOGINEDNAFTKGLFNAAKDTKTNAAVTVGDLNAVAQTPLTFAGDGTGTTAKKL 1453  
Db 1398 vndankkginednafvkgleaksadnktknaavtvgdlnavagtpltfagdglttakkl 1457  
QY 1454 GETLTIKGGQTDNKLTDNNIGVWAGTDGFTVKLAKDLTLNLSNVNAGGTRIDEGKISFVD 1513  
Db 1458 getltikggqtdnkltdnnigvvgagtdgftvklakdltnlnsvnaggtkiddkgsfvfd 1517  
QY 1514 ANGOAKANTPVLNSANGLDLGGKRIISNCAAVDDNDVNFQFNEVAKTVNNLNQNSGA 1573  
Db 1518 ssgqakantpvlansangldl----- 1536  
QY 1574 SLPFVVTDANGKPIGTGCKPQKAIKGADGKYVHANANGVPVDKDGKRPITDADKLANLAA 1633  
Db 1537 ----- 1536  
QY 1634 HGKPLDAGHOVVASLGGSDAITLTNIKSTLPQIDPTNTGNANAGQAQSLPSLAAQOSN 1693  
Db 1537 ----- 1536

QY 1694 AASVKDVLNWFNLOTNHNQVDFVKAYDTVNFVNGTGADITSVRSADGTMSTNVTNTALA 1753  
Db 1537 ----- 1536  
QY 1754 ATDDGDNVLKAKDGKFKYKADDLMPNGSLKAGKSADAKTPTGTLVNPNAKSGTGDAV 1813  
Db 1537 ----- 1536  
QY 1814 ALNLSKAVFKSKDGTTTTSSDGSISIQGKDNSSITLSKDLNVGGKVISVNGGKTDP 1873  
Db 1537 -----ggkvisnvvgktkt 1551  
QY 1874 DAANVOOLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSSNRTVIKAGTVLGKG 1933  
Db 1552 daanvqqlnevrnllglgnagnndnagvniadiKKdpnsessnrtvikagtv199kg 1611  
QY 1934 NNDTEKLATGVOGVVDKGNANGDLSNVWVKTKQDKGSKALLATYNAAGQTNLYLTNPA 1993  
Db 1612 nndteklatgvgvvdkgndgnangdlSNVWVKTKQDKGSKALLatynaagqtnlylcnpa 1671  
QY 1994 EADIRINEQIRFFHVNDGNOEPVVGQRNIDSSASGKHSVAIGFQAKADGEAAVAIGRQ 2053  
Db 1672 ealdrineqgIrffhvnndgnqepvvgqrngldsaagkhsvaigfqakadgeaava19frq 1731  
QY 2054 TOAGNQSTATCDNAQATGDOSIATGTGNVVTGKHSAGIDPSTVKADNSYSVGNNNQFID 2113  
Db 1732 tqagngslagdnagaCgdgslalgtgnvvagkhsagaigdpstvckadnsysvgnnnqftd 1791  
QY 2114 ATQTDVFGVGNNTVTSNSVALGSNSAISAGTHAGTQAKKSDGTAGTTTATAGTVK 2173  
Db 1792 atqtdvfgvgnntvtsnsvalgsnsaisagthagtqakksdgtagttttatagtvkg 1851  
QY 2174 FAGQTAVGANSVGASGAERRIONVAAGEVSATSDAVNGSOLYKATOGIANATNELDHRI 2233  
Db 1852 fagqtavgavsvgasgaerrionvaagevsatSDAVNGSOLYkatogianatneldhri 1911  
QY 2234 HONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNCOCAVAVGLSKLSDNGQWFK 2293  
Db 1912 hqenenkanagIssamamasmpqayIpgRsmvtggIathngggavavglsklSDngqwwfk 1971  
QY 2294 INGSADTOGHVGAAGVAGFHF 2314  
Db 1972 Ingsadtghvgaavgagfhf 1992  
RESULT 7  
AAR99393 ID AAR99393 standard; Protein; 2353 AA.  
AC AAR99393;  
XX 15-JAN-1997 (first entry)  
DT Haemophilus adhesion protein HA2.  
XX Haemophilus adhesion protein; HA2; hsf protein; vaccine.  
DE Haemophilus influenzae type b strain C54.  
KW WO9630519-A1.  
OS 03-OCT-1996.  
PN 22-MAR-1996; 96WO-US04031.  
XX 24-MAR-1995; 95US-0409995.  
XX (UYSL-) UNIV ST LOUIS.  
PA (UNIW ) UNIV WASHINGTON.  
XX Barenkamp SJ, St Gene JW;  
XX WPI; 1996-455364/45.  
DR



	QY	1493	NLSVNVAGGTRIDEKGISFVDANGQAOKANTPVLSANGLDL---GGKRITSNICAAVDDNDA	1549
	DDB	1599	n-----ngvltietdgdegnglvtakavidavnkagwrvtktgqngqd--	1644
	QY	1550	VNFQOFNEVAKTVMNLNQNSNGASLPVFVTDANGKPINGTQDKPOKAIGA-DCKYYHA	1608
	DDB	1645	----fatvasgt------vtfadg---ngtcaevtkandgsitkvnyk	1681
	QY	1609	NANGVPVDKGKPTTDADKLANLAHGK---PLDAGHQVVVASLGNSDAITLTNIKSTLP	1665
	DDB	1682	vadglklidgd-kivadtvtlt--vadgkvctapmgdggkfvdasgladalnklstwtat--	1736
	QY	1666	QIDTPNTGNANGAOQLPSLSLAAQQSNAASKVDVLNWGFNIQTNNHQNQVDF-----VK	1718
	DDB	1737	-----aqkegtgevd--pansagvekadg-kvtfkagdnlikqsgkdfcyslkkelk	1787
	QY	1719	AYDTVNFPVN---CTGCADTVRSADGMSNITV---NTALAAATDDDNVLIRAKDGKFVK	1772
	DDB	1788	dltsefkdangrgtgestki-----tkdgltitpangagaagantantsvtkgd----	1838
	QY	1773	ADDLMPNGSLKAGKSASDAKTPTGLSLVNPNAKGSTGDVALNLSKAVFKS-----K	1826
	DDB	1839	-----isagnka-----vtnvsvglkkfgdghtlangtvadfekhydynayk	1879
	QY	1827	DGTTTTVTSSDGISIQGKDNSSIITLSKDLNVG-----GKVIS---NVCGTKGTDDAANV	1878
	DDB	1880	dit-----nldekgaadnptvadhtaavtvgdlrglvwvisadkttdepqeqyn-aqv	1930
	QY	1879	QOLNEVRNLLGUG-NAGMDNADGQVNIADTIKKDP-----NSGSSNRRTVTIKAGTV	1928
	DDB	1931	rnanevkfksgnglnvsaktlingrvitfelakgevvnksneftvkadgsenlvkvgdm	1990
	QY	1929	LGGKGNDD--TEKLATGVGVGVQDKGNANGDLSNVWVKTKQDGSKALLATYNAAQTIN	1986
	DDB	1991	yyskedidpatstkpmtg---ktekykvengkv-----vsangsktevtltnksg---	2037
	QY	1987	YLTNPNPAEATDRINEQGRIFRPHVNDGNQEPVVQGRNGIDSASGKHSHVATGFOAK---	2041
	DDB	2038	yvcgn--qvadaiahsgef-----gladaaeakaesaakdkqlskd	2079
	QY	2042	-----ADGEAAVATRQRQAQNOIAICDNAQATGDQSITAIGTNVVTGKHSGAIGDPS	2095
	DDB	2080	kaetvnaahdkvrflanglntkv---seatvestdangdk-----vtt-----t	2118
	QY	2096	TVRADNSYSVGNNGQPIDATQTDVFGV---GNIIIVT-----ESNSVALGSNSAISAG-	2145
	DDB	2119	fvtcd-----velplqiyntdangnkikvkadgkwylmadgtsnkevltlgn	2167
	QY	2146	--THAQTOAKSKSDGTAGTTTATAGATGVKVGFAQTAVCAVSVG-----ASGAE	2191
	DDB	2168	vdangkkvvkvtengadkwyyfnadgaadktkgevsndkvtadekhvrlidpnqnsgkg	2227
	QY	2192	RRTQNYAAGEVSATSTDVANGSQLYKATOGIAN---ATNELDHRLIHONENKANAGISSAM	2248
	DDB	2228	vvldnvangelsatscdalngsqlyavakgvctnlaggvnnlegkvknvgkradagtasal	2287
	QY	2249	AMASMPQAYIPGRSMYVTGGTATRHNGQAVAYAGLSKLSLDNGQWFKINGSADTOGHGAUV	2308
	DDB	2288	aasqlpqatmpgmksmvaigssyvgqnglaigvrsidsdngkviirlsgttcnsgktgvaa	2347
	QY	2309	GAGPHF 2314	
	DDB	2348	gvgyqw 2353	
	RESULT	8		
	AAAB23860			
	ID	AAAB23860 standard; Protein; 2411 AA.		
	XX	AAAB23860;		
	XX	17-JAN-2001 (first entry)		
	DT			

Db 227 -esvdlvsayn--nvfeitgdkntl-----dvvlatakengkttevkftpktsvikekgkl 279  
Qy 210 AYAKGFANAFGRSTAGNYSVLAUGL-TAKA-----EKGYTI-AIGSNAQAIFYGALA 261  
Db 280 ftkennndtnkvsntatdntdegvlvtakavldavnkagvrwtktttangqgdftatva 339  
Qy 262 LGADTRVLDLYGTALGVSQILNNNNN--NNKAYVPEGNGSNIKSSK-----AT 309  
Db 340 sg--tnvtfesgd--gttasvtdktngngltvkydakvgdglkfdskkivadtaltvt 395  
Qy 310 GNGLFSGSTIRKRIINVGAGVEDTDAVNVQOLKAVENL---AKROIIFKGDONGGVK 366  
Db 396 ggkvaelaekeddkklvnag-----dlvt-----alignswkakaeadtdtdgalegis 444  
Qy 367 K----KLGETLTIKGE-----TOADKLTDDNNNIGVYVDDNNTGLVKYLAKNL 409  
Db 445 kdvkagagvtvfkagknlkvkqdganfyslqdalgtlsitlgttgngngndaktvink 504  
Qy 410 SGLE-----TVSTKNLTASEKVTVGSGNNTAELQSGGL-----TFPTTNASTDKT 455  
Db 505 dgltitpagnngtvtgntisvt-kdglkagnkaltvnasglrayddanfvlvlnnsatdl 563  
Qy 456 VYGTGGLKFTDNSN-----TALEDTRITKD---KIGF--SNKAGTVBENKPYLBDK 503  
Db 564 rhvedaykgllnlnneknankpivtdstaatvgdlrlglwvstknktkeesnykqgade 623  
Qy 504 L----KVGNSLT----NNGGJTVNNTIGGSKQIQVGADG---IKPADVNVVNSNAKFGT- 553  
Db 624 vltfgagaatvtsksengkhtitvsvaetkadsglekdgdtkiklvdnqntdnlvtvgnn 683  
Qy 554 -TRITE-----EEGFPADAD-GKVDKSPYLDKKQLQVGGVTKTKDSINAGDOKISNVK 606  
Db 684 gtavtkggtvtgtatdargkv-----tvkdatanadkdvkvatv 725  
Qy 607 D-ATDDTDAVYKQKQVQODAGALQSFSSIRDEKQGEFTISNLSNG---NTPNTEFIT 663  
Db 726 dvatainsaafvktenltside-----dnptdngkdalkagdtlt 768  
Qy 664 F-AGEN-----GISINDIAKG-KVKV-----GIDPINGLELT-TPKLTVGSDKDG 704  
Db 769 fkaagknlkvrdgknlftldaknlvktakvstltlgtgntpggttatkpvnitstadg 828  
Qy 705 KTQLVIEQVAGSNDTKNI-IRGLSPTL--PSITNAGGVRTTEQGNNTTSDEDSKAASIG 761  
Db 829 ln---faketadaagsknvykglattlttpep-----agaksghvlnvdatk-knsaasle 881  
Qy 762 DILTFENLKNNSVGFVSTYNTVDFIDGNATTAFTYDETNTQTSKVTYDVNVDEKTIE 821  
Db 882 dviragwnigngnnvdvtydtvntfddstgtttv--tqkadkgadvkigaktv 939  
Qy 822 LTGDNKGTNKIGVKTTLTTLTTNANGKATNFSTDDND-----ALVNAKDIAENLNTLAK 874  
Db 940 ikdhngk-----lftgklidkdnngatvsedgdgkdtglvltaktvldavnsksw 989  
Qy 875 EIHTTKTADTALQTFVKVKDGAH-----DDEITVCKDGTONGKTYNTLKLKGENGLTV 929  
Db 990 rvtgagataet-----gacavnagnaetvt-----sgtsvn-----fknagnatta 1029  
Qy 930 ATNRKDG---TVTFGINTQSGKLKAG-----DSTTLNKDGLSINKPA-----SNEQIQVG 974  
Db 1030 tvskdngnlnkvdyvngvgdglkigdkkivadtltltvtgkvsvpagansvnnkklvn 1089  
Qy 975 ADGVK-----FAKVDK----- 985  
Db 1090 aeglatalnlswtakakadyadgesegedqevkagdkvtfkagknlkvkqsekdfcysl 1149  
Qy 986 -----GNSSTGIDCTSR-----ITKQDQIGFTGANGSLDTRKPHLTOKLKVGEVEIT 1032  
Db 1150 qdtlgtitsltlgtangrntdgtvinkdgtltitlangaaagt-----dasngntisvt 1203  
Qy 1033 NTGINAGGKKITNIQSG-----DITQNSNDA-----VTGGRVYDLKTELESKINSAA 1079  
Db 1204 kdglsagnkeitnvksalkkykdcntqtagatqpaantaevakqdlvldtkpatgaagnga 1263

Qy 1080 KTAQNSLHEFSVADEOQHFTVSNPYSSYDTS-----KTSDVITTFAGENGITTKV-- 1129  
Db 1264 dakapdttaatvgdlrglvgwlsakktadettqkdfhaavknanavefvgkngatvsakt 1323  
Qy 1130 -NKGVVRVGIDQTK-----GLTTP-----KLTVCNNNGKGV-IDSKDQGTITGLSN-- 1175  
Db 1324 dnnghkvtidvaeakvgdglektdgkiklvtdntdgnlltvtdatakgsavakgefnv 1383  
Qy 1176 ---TLANVTN-DGAGHALSOG---LANDTKTRAASIGDVLNAGFNLOGNEAVDFVS 1236  
Db 1384 ttdattaggttneargkvvvvkgsgngatataetcdkvvkvtvgdvaka----- 1428  
Qy 1227 TYDVFDFIDGNATTAFTYDGT-----SKTSKVYVYVNVNDKNTIETVSDKKLGVKTTTLT 1281  
Db 1429 -----indaatfvkveddsatisiddspddgandalkagdtltlkagknlkvk----- 1476  
Qy 1282 KTSANGNATFSAADGDALYKASDIATHLNTLAGDIQTAKGASQASSASVVDADGNKVI 1341  
Db 1477 ---rdgknitfalan-dlsvksatvsdkls-----lgtngnkv 1511  
Qy 1342 YDSTDKKYYQVNDKGQVDKNKEVAKDKLVAQAOTPDGTLAQMNKVSINKVEQVNDANKQ 1401  
Db 1512 itsdtkglnfakd-----sktgddanihlnglasiatlttdllnsgattnlngngicdnekr 1567  
Qy 1402 GINEDNAP-----IKGLENAAKDTKTNAAYTVGDLNVAQOTPLTFAGDTGTTAKKLGET 1456  
Db 1568 aasvkdvlngawnvrgvkpasannqveni-----dfvatydtvdfvsgdkdt-----s 1616  
Qy 1457 LTIKGGOTDKNLTDDNIG---VVAGTDG--FTVKLAKDLTLNLSNVNAGTRIDEXGIS 1510  
Db 1617 vtve--skdngkrtevkigaktsvikdhngklftgkelkdann-----ngvt 1661  
Qy 1511 FVDANGAKANTPVLSSANGLDL---GGKRISNIGAAVDDNDVAFKQFNEVAKTVNNLNN 1567  
Db 1662 vtetdkgdegnglvtakavidavnkagrvkttgangqndd-----fatvasgtn--- 1711  
Qy 1568 QSNSGALPFPVVVDANGKPIGTGPKOKAIKGA-DGKYYHANANGVPVVDKGPITDAD 1626  
Db 1712 -----vtfadg---ngttaevtkandgsiltvkynkvvadgklldgd-kivadt 1756  
Qy 1627 KLANLAHAGK---PLDAGHQVVASLGGNSDAITLTNKTSLPQIDTPNTGNANAGQAQSL 1683  
Db 1757 vlt--vadgkvtepnngdkkfvdasgladalnklswat-----agkegtgevd-- 1804  
Qy 1684 PSLSAAQOSNAASVKDVLNVGNFLOTNHNQVDF-----VKAYDTVNFVN---GTGADI 1733  
Db 1805 pansaggevkagd-kvtfkagdnhikikqsgkdfkyslkelkdltsvefkdganggttges 1863  
Qy 1734 TSVRSADGTMSNITV---NTALAAATDDDNVLIKAKDGKFFYKADDLMPNGSLKAGKSASD 1790  
Db 1864 tkf-----tkdgltitpangagagasantantitsvtdkg-----isagnka-- 1903  
Qy 1791 AKTPGLSLVNPNAKGSTGDAVALNLSKAVFKS-----KDGTTTTVSSDGISIQGK 1844  
Db 1904 -----vtvnvsglkkfgdghtlangtvadfehydnaykdlit-----nldeka 1947  
Qy 1845 DNSSITLSKDLNLYG-----GKVIS---NVGKGTDKTDAANVQOLNEVRNLLGLG-NAGN 1895  
Db 1948 dnnptvadntaataatvgdlrglvgwisadkttegepnqeyn-aqvrnanevkfsgnginvsq 2006  
Qy 1896 DNADGNQVNIADIKKDP-----NSGSSNRTVTKAGTVLGGKGNND---TEKLATGG 1944  
Db 2007 ktlngtrvittfelakgevvksneftvnadgsenltkvvgdmyyskedidpatskpmgtg- 2065  
Qy 1945 VQGVVDKDGNGANDLSNVVWTKQDGSKKALLATYNNAGQTNLYLTNNPAEALDRINQGI 2004  
Db 2066 ---ktekyvengkv-----vsangsktevtltnksg---yvtgn--qvadaiaksqf 2111  
Qy 2005 RFFHVNDGNQBPVQGRNGIDSSAGKSHVAIGFOAK-----ADGEAAVAIGRQ 2053  
Db 2112 el-----gladaaeaka faesakdkqlskdkaetvnahdkvrfangln 2155



Thu Sep 13 14:18:07 2001

```

CC influenzae species. A truncated protein has a significantly higher
CC amount of recovery than a full-length protein.
XX
SQ Sequence 1104 AA;

Query Match      8.3%; Score 968; DB 21; Length 1104;
Best Local Similarity 24.6%; Pred. No. 3.3e-35;
Matches 391; Conservative 168; Mismatches 441; Indels 590; Gaps 55;

QY 813 VNVDEKTIELTGDNGKTKKIGKVGKTTTLTT-----NANGKATNFSTDN--- 856
DB 17 vvvseltrahtkcasavavavilatalasacaannntsvtnglnaygd-tnfntnnsia 75
QY 857 DALVNAKDIAEALNTLAKIEHTTKGTADTALQTFVKKDGATDDETIIVG----- 906
DB 76 dlekhvqdaykglllnnek-----dtnksflv-----adntaatvgnrlkgwvls 122
QY 907 -KDGTONGKT-----VNTLKLKGENGLTV-ATNKDGTVTTFGLNTOSGLKAGDSTTLNKG 959
DB 123 skogtrneksyqvkadevlfctgsgaatvsssskdgkhtitlsvckgsfaevktdattlg 182
QY 960 LSIKNPASNEQIOVGADGVKFAKVDKGSSTGIDGTSRITKQDIOGTFGANGSLDTTKPHL 1019
DB 183 -----qvnadrgkv---kaedengadvdkvatvdkvakaındaatfvkves 226
QY 1020 TKDKLKYGEVEITNTGINAGGKKITNIQSGDITQNSDNDVATGGRVYDLKT--ELESKINS 1077
DB 227 tdddieng-----aag-----netcdqalkagdtltlckaglnlkakldq 266
QY 1078 AAKTAQNSLHEFSVADEQGNHFTVSNPYSSTYSKTSIDVTTFAGENGITTKVNGKVVRVG 1137
DB 267 ngkav-----tfalak-----ldvtsakvsdkls-----lgdktnkvd 300
QY 1138 I-DQTKGLTTPKLVGNNGKGIVIDSKDQNTITGLSNTLANVNDGAGHALSOGLAND 1196
DB 301 itsdanglklak--tgngng-----qngnvhngiaastltd-tlgtmtqasngvavq 350
QY 1197 TDKTPAASIGDVLNAGFNLCQGEAVDFSVSTVDYDFIDGNATITAKVYDDTSTKSKVY 1256
DB 351 -nhnraasvadvlnagwnlqngasvdfvnaeydvtvfvngntnvnvtttdtahnkktvrv 409
QY 1257 DVNVDNKITEVTSKDKLGVKTTTLTKTSANGNATKFSNAAGDALVKASDIATHLNTLAGD 1316
DB 410 dv-----e----- 411
QY 1317 IQTAKGASOASSASYVDADGNKVIYDSTDKKYYQVNDKQVDRKNEKVDKLVAAQATP 1376
DB 412 -----tglpvgyvtcdgktvv--kvdnkyyeakdgdsadmdkkv-----e 449
QY 1377 DGTLAGQMNK--SVINKEQVNDANKKQGINEDNAF-----IKGLENNAKDRTKNAAVTVG 1430
DB 450 ngelaktkvklvsasgqnpvklsnvaeagteendavsfkqlkalgekvtllcasnayang 509
QY 1431 DLNAVAQTPLTAGDPTGTTAKKLGELTTLKGGQTDNTNLNIGVAVAGTGFTVKLAKD 1490
DB 510 n-----dadggkatqlnnglnfkfstgdgell---nikvendtvtftpk----- 551
QY 1491 LTNLSVNAAGTRIDEKIGISFVDANGQAKANPVLISANGLDGGKRISNIGAIVDNDVAV 1550
DB 552 -----kgsqvgedgkatlqngtktdgl----- 575
QY 1551 NFQKFNEVAKVTNNLNQNSGASLFFVVTDANGKPIGTGDKPQKAIKAGDGKYYHANA 1610
DB 576 --veaseiveslnklgw-----gasnetlvksqdkvt----- 591
QY 1611 NGVPVDKGRPIPTDADKLANLAHAKPLDAGHQVVASLGSNSDAITLTNFKSLPQIDPP 1670
DB 592 --vgydkdggeld----- 618
QY 1671 NTGNANAGQASLPSLSAAQASNAASVKDVLNVGFNLNHNQVDFVAYDTVNFVNGTG 1730
DB 619 -----lkaenlkvkq-----dgnftyaik 639
```

```

QY 2054 TQAGNQSTAIQNAQATQDQSIAGTGNVYTKHSGAIGDPSTVKADNSYSVGNNOQFID 2113
DB 2156 tkv---saatvestdangk-----vtt-----tfvkt-----ve 2183
QY 2114 NTQDVEFGV---GNNTITVT-----ESNSVALGSNSAISAG---THAGTQAKKSDGTAGT 2161
DB 2184 lplqtqyntdangnkivkdkagkywelnadgtasnkevltlgnvdangkkvkvvtengadk 2243
QY 2162 TTTAGATGVKGFAGTAVGAVSVG-----ASGAERRIONVAGEVSATSTDA 2209
DB 2244 wyytnadgaadtkegsndkvstdekhhvrlpnnngngkgyvldnvnangeisatstda 2303
QY 2210 VNGSOLYKATOGIAN---ATNELDHRIRHONENKANAGASSAMAMASMPQAVIPGRSMVTG 2266
DB 2304 lngsqlyavakvntlagvnnlegkvnkvgradagtasalaasqlpqatmpgksmval 2363
QY 2267 GIATHNGQGAVALGSLKSDNGQWFKINGSADTQGHVGAAGAGFHF 2314
DB 2364 agssyggqnglaigvarisdngkvliirlsgtttsqgktgvaagvgqvw 2411
```

```

RESULT 9
AAB23856
ID AAB23856 standard; Protein; 1104 AA.
AC
AA
DT 17-JAN-2001 (first entry)
XX
XX Haemophilus influenzae adhesin (Hia) protein from NTHi strain 29.
DE
XX Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;
KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;
KW diagnosis; immunogenic; antigen.
XX
OS Haemophilus influenzae.
XX
XX WO200055191-A2.
XX
XX 21-SEP-2000.
XX
XX 16-MAR-2000; 2000WO-CA00289.
XX
XX 16-MAR-1999; 99US-0268347.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Yang Y, Klein MH;
XX
XX WPI: 2000-618897/59.
XX
XX N-PSDB; AAN92495.
XX
XX Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
XX use as antigens and vaccines and for treating Hemophilus influenzae
XX infection.
XX
XX Claim 1: Fig 20; 275pp; English.
XX
XX The present sequence represents a Haemophilus influenzae adhesin (Hia)
XX protein from the non-typeable Haemophilus influenzae (NTHi) strain 29.
XX Hia genes and proteins have antiinflammatory, auditory and antibacterial
XX activities, and can be used in the production of a vaccine. An
XX immunogenic composition comprising an Hia gene, a polypeptide encoded
XX by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
XX protection against disease caused by Haemophilus strains in a
XX susceptible host, preferably a human. An Hia protein is useful as an
XX antigen, in immunogenic preparations including vaccines, as a carrier
XX for other immunogens, and in the generation of diagnostic reagents. Hia
XX is useful for treating diseases caused by the infection of Haemophilus
XX influenzae such as meningitis, epiglottitis, septicemia and otitis
XX media. Recombinant production of Hia favours high recovery of the
XX protein compared to the low recovery of native protein from Haemophilus
```



Db 412 -----tglpqvvtedgktvv--kvdnkyyeakqdgsgadmdkkv-----e 449  
Qy 1377 DGTLAQMNVR--SVINKEQVNDANKKQGINEDNAF----IKGLENAKADTKTKNAAVTVG 1430  
Db 450 ngelaktkvklysaagqnpvkslnvaegteendavsfkqkalqekqyvtltasnayangg 509  
Qy 1431 DLNVAOTPLTITAGDTGTTAKKLTGTLIKGGQTDNKLTDNNIGVAVAGTDFVTKIAKD 1490  
Db 510 n-----dadggkatqllnglnfkfstgdgell--nnikvendtvtfepk---- 551  
Qy 1491 LTNLSVNVAGTRIDEKISFISFVANGQAKANTPVLUSANGLDGGKRSINIGAAVDDNDV 1550  
Db 552 -----kgsqvqgedgkatigngtkttdgl----- 575  
Qy 1551 NFQPFNEVAKTVNNLNQNSGASLPFFVTDANGKPIINGTDGKPKQAIKAGDKRYIHANA 1610  
Db 576 --veaselvesinklgwk----- 591  
Qy 1611 NGVPVDKDGKPIITADKLANLAHAKPLDAGHQVVASLGGNSDAITLNLINIKSTLPQIDTP 1670  
Db 592 --vgvdkdsgeld-----gasnetlvksgdkvt----- 618  
Qy 1671 NTGNANAGQASLPISLAAQSNAAASVKDLNVNGENLQTNHNOVDVFKAYDVTNFEVNGTG 1730  
Db 619 -----lkagenlkvkq-----dgtntfyalk 639  
Qy 1731 ADITSVRSADGTMSNITNTALAAATDDGNVLKAKDKGFYKADDLMPNGSLKAGKSASD 1790  
Db 640 delgkvks-----vefkd-----tangst-----ngast 662  
Qy 1791 AKTPTGLSLVNPAGKSGTGDVAVALNLSKAVFKSKDGTITTTVTSDDGISIOCKNDSST 1850  
Db 663 kitdgtitsangangaa-----atdaeklk 689  
Qy 1851 LSKDGLNVGKRVISNVGKGTGTDAAVVOQLNEVRNLL-----GLGNAGNDNADGN 1901  
Db 690 vasdgisagnkavknvvgslkfgdanfnlptasadnltkqydaykgtlnidekqadkq 749  
Qy 1902 QVNIADIKKDPNSSSSSNRTVIGKAGTVLGGKGNNDTEKLATGGVGVGVDKDNANGDLSN 1961  
Db 750 tiltvad-----nt-----datv-----gdrlg 766  
Qy 1962 V-WKTKQDKGSKALLATYNAAGQNTVLTNNPABADRINEQGLRFFHVNDGQEPVVOG 2020  
Db 767 lgwv-isadkttegeinkeyna-----qvrnanefvfksg 799  
Qy 2021 RNCIDSS---ASGKHSVAIGQAKADGEAAVAIGRQTOAGNOSIAIGDQAATGDQSIAT 2077  
Db 800 -ngihvsgktvngrrreit--felakdena-----lafgygskalrtdntvai 842  
Qy 2078 GTGNVVTGKHSIGAIDPSTV--KADNSYSVGNNOFIDATQTDVFGVGNV----- 2125  
Db 843 gtgnvvnnaeksgafgdpnyiedkaggsyafgndnr-itskntfvlngngvnaakykangdvd 901  
Qy 2126 -----ITV-----TESNVALGNSAITSAGTHAGTOAKKSDGTAGTITTT 2164  
Db 902 tetvtvkdkgkettvtvpkalgatvensvylgnk---statdkgknlksdgtagnttt 958  
Qy 2165 AGATGVKGFAGQATVAGVSVGSAEERRIQNVAAGEVSATSTDAVNGSOLYKATQGTIAN 2224  
Db 959 agtgtvngfagatahagvsvgaasgeerriqnvaageisatstdaingsqlyavakgvtn 1018  
Qy 2225 ATNELDHRHONENKANAGISSMAMASMPQAYIPCRSMVVTGGIATHNGOGAVAVGLSKL 2284  
Db 1019 ----legqvnkvkradagtasalaasqllpqasmpgksmvsivsiagssyagqnglaigvsi 1074  
Qy 2285 SDNGOWFEKINSADTQGHVGAAGVAGPHEF 2314  
Db 1075 sdngkviirigsntnsgkgtgvaagvggq 1104

RESULT 11  
AAB23857

AA023857 standard; Protein; 1004 AA.  
AAB23857;  
17-JAN-2001 (first entry)  
Haemophilus influenzae adhesin (Hia) protein from NTHi strain M407.  
Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;  
non-typeable Haemophilus influenzae; antinflammatory; auditory;  
antibacterial; meningitis; epiglottitis; septicaemia; otitis media;  
diagnosis; immunogenic; antigen.  
Haemophilus influenzae.  
WO2000055191-A2.  
21-SEP-2000.  
16-MAR-2000; 2000WO-CA00289.  
16-MAR-1999; 99US-0268347.  
(CONN-) CONNAUGHT LAB LTD.  
Loosmore SM, Yang Y, Klein MH;  
WPI; 2000-618897/59.  
N-PSDB; AAA92496.  
Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for  
use as antigens and vaccines and for treating Hemophilus influenzae  
infection  
Claim 1; Fig 21; 275pp; English.  
The present sequence represents a Haemophilus influenzae adhesin (Hia)  
protein from the non-typeable Haemophilus influenzae (NTHi) strain M407.  
Hia genes and proteins have antinflammatory, auditory and antibacterial  
activities, and can be used in the production of a vaccine. An  
immunogenic composition comprising an Hia gene, a polypeptide encoded  
by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
protection against disease caused by Haemophilus strains in a  
susceptible host, preferably a human. An Hia protein is useful as an  
antigen, in immunogenic preparations including vaccines, as a carrier  
for other immunogens, and in the generation of diagnostic reagents. Hia  
is useful for treating diseases caused by the infection of Haemophilus  
influenzae such as meningitis, epiglottitis, septicaemia and otitis  
media. Recombinant production of Hia favours high recovery of the  
protein compared to the low recovery of native protein from Haemophilus  
influenzae species. A truncated protein has a significantly higher  
amount of recovery than a full-length protein.  
Sequence 1004 AA;

Query Match 8.2%; Score 959.5; DB 21; Length 1004;  
Best Local Similarity 23.7%; Pred. No. 7, 1e-35;  
Matches 373; Conservative 168; Mismatches 336; Indels 697; Gaps 57;

Qy 819 TIELTGDKNGTKNIGVKYKTTTLTTTANGKATNFSTTD-----NDALVNAKOAENLN 870  
Db 46 tvqatttggtstnglkayg-stnnpnfnaagsatdlarqfdgaydglnlnekdan 104  
Qy 871 TLAKETHTTKGTADTALQTFVKVKDGTDEITVTC-----KDGTONKGT---- 915  
Db 105 ll-----vtdokaatvgnlrlklgwvlskngtrneksqqvk 140  
Qy 916 -VNTLKLKGENGLTVATNKDG---TWTFGINTOSGLKAGDSTTLNKGDSINPASNEOI 971  
Db 141 hadevlfeqkdgvvtksengkhtvtf-----tlekd-invknatvskl 185  
Qy 972 QVGADGVKFAKVDKGNSSGTIDGTSRTTKDQIGFTGANGSLDITTKPHLTKDKLVGEVEI 1031



CC media. Recombinant production of H1a favours high recovery of the  
CC protein compared to the low recovery of native protein from *Haemophilus*  
CC influenzae species. A truncated protein has a significantly higher  
CC amount of recovery than a full-length protein.

AA	Sequence	1002 AA;
SQ		

Query Match	7.7%;	Score 895.5;	DB 21;	Length 1002;
Best Local Similarity	24.9%;	Pred. No. 4.7e-32;		
Matches 345;	Conservative 159;	Mismatches 352;	Indels 527;	Gaps 53;

[illegible]

Db	610	fgdanfnplt-----ssadnltkqydnaykgltnldekakgkqtpvadnt---aat	658
Qy	1928	VLGKGKGNNDTEKLTATGGVGVGDKNAGCDLSNV--WV--KTKQDGSKALLATTYNAAGQ	1984
Db	659	v-----gdrlrglgwvisadktlgeakeysaqvrnane	690
Qy	1985	TNYLTNPAPAEADRINPEQGRFFHFVNDGNQPEVVQVRNGIDSSAGKHSVAIGFOAKADG	2044
Db	691	vkfksnginvsqktdlndgrt-----eltfelakde	721
Qy	2045	EAAVAIGRQTAGNQSTAIAGNQAATGDSOIAIGTGNVVTGKHSAGIDPSTV--KADNS	2102
Db	722	na-----lafgsgkalrldntvalgtgnvnaeksgafgdpnyiedkaggs	767
Qy	2103	YSGVNNNQFIDATQTDVFGVGNN-----ITV-----TES	2131
Db	768	yafgdnr--ltskntfvlvgvnaakyangdvdtetvtvkdkdgtetvtvpkalgatve	826
Qy	2132	NSVALGNSAISAGTHAGTQAKKSDGTAGTTTTAGATGTVKVGFAQTAVGAVSVGASGAE	2191
Db	827	nsvylgnk---statkdkgnlksdgtagntttagtftgtnvgfagatahgvavsgaqsee	883
Qy	2192	BRIGNVAAGEVSATSTDVNGSGLYKATOGIANATNDELDRHTIQHENKANAGISSAMAMA	2251
Db	884	rriqnvaagealsatstdalingsqlayavakgvtn----lagqvnykvgrkadagtasalaas	939
Qy	2252	SMPQAYIPGRSMVTGGIATHNGOGAVAGLSKLSDNQGWVFKINGSADTOGHGVGAACVAG	2311
Db	940	qlpqasmgsksmvslagsgyqgggtaigvrslsdngkvlrlfsgcltnsqgktgvaaqgv	999
Qy	2312	FHF 2314	
Db	1000	vaw 1002	

RESULT 13  
AAW56322

AAW50322  
ID AAW56322 standard; Protein; 2039 AA.

AA AAW56322;

19-AUG-1998 (first entry)

DE Haemophilus paragallinarum antigenic protein #2  
XX

XX	Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis; vaccine; chicken infectious coryza; CrC; fowl.
KW	
XX	

OS Haemophilus paracollinarum.

XX	Key	Location/Qualifiers
FH		

Peptide	1..70	/label = signal
FT		
FT		

FT	Protein	71..2039
FT	Protein	71..2039

xx

PN XX  
W09812331-AL.

PD 26-MAR-1998.  
XX

PF 12-SEP-1997; 97WO-JP03222.

PR 19-SEP-1996; 96JP-0271408.

PA (KAGA) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.  
PA (KAGA) CHEMO-SERO-THERAPEUTIC RES. INST.

XX	Hamada E	Matsuo Y	Sakaguchi M	Moriuchi C	Makino S
DI					

XX  
XX

DR N-PSDB; AAV22837.

XY



Thu Sep 13 14:18:07 2001

us-09-361-619-11.rag

```
Db 1585 dgidagnkksinvadgd-----isptegdvvtgr-----qlyalmqk--- 1621
Qy 1696 SVKQVLNVGNFNLQNHQVDFVKAYDTVNFVNGNGADITSVRSADGTMSNITVNTALAAT 1755
Db 1622 -----giryvdevsptkqtaptastcggattantagg----- 1657
Qy 1756 DDDGNVLIAKADGKFKYKADDLMPNGSLKAGKSASDAKPTPTGLSLVNPNAKGSTGDAVAL 1815
Db 1658 -----vapagnvatgdia-----ptqpal--pemktalvgdhlav 1690
Qy 1816 NNLXAVFKSKDGTFTTTVSSD---GISIQKDNSSITLSKDKGLNVGKVKVSNVGKCTKD 1872
Db 1691 -plgslkihgdmvkttisagnqvglq--pnlsi---ennlvignkpeklaaqae 1744
Qy 1873 TDAANVQQLNEVRLLGLGNAGN--DNADGNQVNIADIKKDPNS---GSSSNRTVIKAGTV 1928
Db 1745 gnalvtnkdd-----gnaamvfnncknmvlvsdkkakpravldgngngaltlv----- 1792
Qy 1929 LGGKGNNDTEKATGGVQVVDKDNANGDLSNVWVKTKDGSKKALLATVYNAAGOTNYL 1988
Db 1793 ----gnddsq--vtlsskkgkdidgn--dlslsvtttert-----nadgglekv 1833
Qy 1989 TNNPAAIDRINEOGLRPFHVNDNQEPVVOGRNGIDSSASKHSVAIGFOAKADGEAAV 2048
Db 1834 etsfaumdd-----glkf-----kadgdkvi 1854
Qy 2049 AIGROTQAGNQSTAGTGDNAQATGDOSIAIGTCNVVTGKHSGAIGDPSTVKADNSYSVGN 2108
Db 1855 -----nkinetvei-----vgdenvtt-----siddnkvkvslnkklaid 1891
Qy 2109 NQFIDATQDVFVGNNITVTESNVALGSNAISAGTHAGTQAKSKDGTAGTTTTAGAT 2168
Db 1892 evk1ptdpda-qkgdsivi-----nnglthagnkvitgvskaad----- 1929
Qy 2169 GTVKGFAGOTAVGAVSVGASGABERRIQNVAAAGEVSATSTDVAVNGSOLYKATOGIANATNE 2228
Db 1930 -----dptsavnrgqintvidnvgnnfnq 1953
Qy 2229 LDHRIHONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGOGAVAGLSKLSLSDNG 2288
Db 1954 vndrigdltesraglagematasignvalpgkttisvgtatfkgenavaimgsrlsdng 2013
Qy 2289 QWVEFKINGSADTQGHVGAAYVAGGFHF 2314
Db 2014 kvgirlsgmstngdkgaamsvgttf 2039

RESULT 14
AAW56319 standard; Protein; 2042 AA.
XX AC AAW56319;
XX DT 19-AUG-1998 (first entry)
XX DE Haemophilus paragallinarum antigenic protein #1.
XX KW Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;
XX KW vaccine; chicken infectious coryza; CIC; fowl.
XX OS Haemophilus paragallinarum.
XX PH Key Location/Qualifiers
XX FT Peptide 1..70
XX FT /label= signal
XX FT Protein 71..2042
XX FT /note= "antigenic protein"
XX PN W09812331-A1.
XX PD 26-MAR-1998.
XX PF 12-SEP-1997; 97WO-Jp03222.
```

```
XX 19-SEP-1996; 96JP-0271408.
XX (KAGA ) 2H KAGAKU & KESSEI RYOHO KENKYUSHO.
XX (KAGA ) CHEMO-SERO-THERAPEUTIC RES INST.
XX Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;
XX WPI; 1998-230318/20.
XX N-PSDB; AAV22834.
XX Antigenic polypeptide from Haemophilus paragallinarum induces HI
XX antibody production - and is useful for diagnosis of and preparation
XX of vaccines for chicken infectious coryza
XX Claim 1; Page 51-69; 108pp; Japanese.
XX The present sequence represents an antigenic protein derived from
XX Haemophilus paragallinarum strain A-221. The antigenic protein
XX stimulates the production of HI antibodies in fowl. The protein
XX and DNA coding for it can be used in the preparation of vaccines
XX for the prevention of chicken infectious coryza (CIC). The protein
XX and its antibodies can be used in the diagnosis and treatment of CIC.
XX Sequence 2042 AA;
XX SQ

Query Match 6.8%; Score 796; DB 19; Length 2042;
Best Local Similarity 20.9%; Pred. NO. 2.7e-27;
Matches 556; Conservative 322; Mismatches 816; Indels 968; Gaps 118;

Qy 1 MNHYIKVIFNKATGTFMAVAEAKS-HSGSSSTAGOVGSSPVI-RLTRVATLAILVIG 58
Db 1 mnkvfkivsvkqgmivvselannkktasqknthttaffqlftkctylallnialg 60
Qy 59 ATL-----NGSAYAQNNS-----KTAFGTGTGNNDN 83
Db 61 aslfpqlanakwlevyassvklstvsqsnvlnpsgaesvgtinspgvgaigvgtndr 120
Qy 84 ASASNEASIAIGSLAKAHANQAIAG-GSKP----- 113
Db 121 -satgialvgvgnktnlakdsialgvgaknestapsvltgkqlarfkelsivmgnay 179
Qy 114 ---DPRNQAANKAGASHAKGK-----ESTAIIGDVLAEGDASIAI 150
Db 180 tqldprgtsketrqgsvvigenaksagqsvslgnsksktnsisigagtfaegkssai 239
Qy 151 GSDDLYLDNRNSTNSKYP-----NGLLSTLIQN-----HTVLRQIRDSNGSQ----- 191
Db 240 gtdk--isgtkyndklpatawng-tgtvpknsiwdifselymgkqtngrdydtttrdpnk 296
Qy 192 -----KY-----RRTAAEGHASTAVGAMAYAKGHFANA 219
Db 297 peafykfsdfkgykvyntptaaptiyagkigalgalgertlaaagemstavgslafaladrsta 356
Qy 220 FGTRS-----TAEGNY-----SLAVGLTAKAKGYTIAIGSNAQAINYGALALG 263
Db 357 mglrsfvakdvvgtaigeestrfakdsvaigntkeasnagmaygykakavvggalaig 416
Qy 264 -----ADPRVDLDYGIAGLGYGSQILNNNNNN----- 289
Db 417 tevaagakfnshqtnllqdnnaayatlnnadsddtktnaltvtqsfdomltnpliv 476
Qy 290 -NNKAYVPECGNGSNIK-----SSKATGNGLFGISGSTITKRIINVG-AGYEDTDAVNVAAQ 342
Db 477 senetylttsagaikktatttdsasgggknaialgsktfaskansvalasayaladaqnafa 536
Qy 343 LKA---VENLAKRQITFKGDDNGTGKVKKLGELTLTKGGETQADKLTONNNNIGVVYDNNNT 399
Db 537 lgsysfvessatntiti-----gvgsyak-gknsfl--ggtwastlsdrt---vvlgnst 585
Qy 400 GLKVKLAKNLGLSETVSTKNTLTASEKVTVGSGNNNTAELOSGLTFTPTPTNA-STDKTVYG 458
```



Db 586 sis-----sg-----sqnalai-gvrvfigndsaslsalngmgstiakeakspdsiaig 632  
QY 459 TDG---LKFTDNGN---TALEDTTRITKDKIGFSNKAAGTVDENKPYLDKDKLVG-NST 510  
Db 633 kearidakdtngtlyqpyvdyetr-----afn-----fnessdym-rqamalgfnak 681  
QY 511 LNNGLTVNNTIGSGNKQIOVGADGTFADVNVNVSNAKFGT-TRITEEIGFADADGK 569  
Db 682 vsrg---vgkmetgin-smaiga---yaqatignstalgvsktdytwegl-----725  
QY 570 VDKKSPYLDKKQLQV-----GGVKITKDSGINAGDQKISNVKDATDQDVAVTYKQLQV 623  
Db 726 --eudpwvsegaisiptsgktvisv---gskgserflvnlasgsdtdavnaqlktv 779  
QY 624 QODADGALQSFISDERKQGFETISNLSNGNTPN--TFETITFAGENGISISNDIAKRG- 680  
Db 780 ee-----rflseilnllngggvkvylsvektningqsg-rvasqirngen 822  
QY 681 ----VKVGIDPINGLTTPKLTVGSDKDGKTQOLVIEQVASGNDTKNIIRIGLSPITNA 736  
Db 823 yeryvkl-----ktqilly-----ldargklngekfdqnsi---852  
QY 737 GGVRTTEQGNITISDEPKSKAASIGDILLNTGFNLKNSNSVGFVSTYNTVDFIDGNATTA 796  
Db 853 -----nkiravvgeleaeysge-----lktasalnqvataq-----leqevitn 891  
QY 797 KVTYDETNOTSKVTYDVNVDEKTIETLTDGNGKTNKIGVKTITLTITNANGKATNFSTDN 856  
Db 892 --nfdkfnq-----yktqienasnad-sarnvggltg 920  
QY 857 DALVNAKDIAENLNTLAKEIHTTKGTADTALQTFKVKKDGATDDETTTVGDKGTQNGKV 916  
Db 921 qiaaqlkannnylndgak-----gqdsiafgwqaktsgannng---lagkqalagfqa 970  
QY 917 NTLKLGKENGLTAVTKNDGVTFTGINTQSGLKAGDSPTLANKDGLSINKPASNQIQVGAD 976  
Db 971 ns-----saenaislgtasdtmt-----gavaigkgatvtagg-----kpsiallgqd 1013  
QY 977 GVKFAKVDKNGSSITGIDTSRITKDKQIGFTGANGSLDITTKPHLTKDKLVGEVEITWTGI 1036  
Db 1014 stvansaistsspmg-----lifnnfagspet-----lgvlisgtagr 1054  
QY 1037 NAGGKKTITNQSODITQNSNDNAVTVGGRY-----DLKTELESKIN-----1076  
Db 1055 e---rklvlnvaagdvsgasteaingsqlyatnfmllskvadgsksnfgnvnlgtdgtitf 1111  
QY 1077 -SAAKTAQNSLHE-----FSVAD-----EOGNHFTVS--NPYSSYDTS 1111  
Db 1112 tn:ig99tqatihdainnvltkglylkadqndptnggqkvvelgnaltisatnqwanngvn 1171  
QY 1112 -KTSDVITFAGENGITTKVKNKVVRVIGIDQTKGITTTPKLTGVNNGNGKGIVIDSKDGQNTI 1170  
Db 1172 yktnnlitynsqn-----gtlilgmredp--svkqitagtynttg-----dannk 1214  
QY 1171 TGLSNTLIANTNCGAGHALSQGLANDTDKTPRAASIGDVLNAGNLOQNGEAVDFVSTYDT 1230  
Db 1215 nqlnntliqqtileatg-----itssvgstnyagfslga-----ds 1249  
QY 1231 VDFIDGNATTAFTYDDTSKTSKVVDVNVNDKNTIEVTSDDKLGCVKTTTLTKTSANGNAT 1290  
Db 1250 vtfsg99agcvkks-----gvsdat-----1269  
QY 1291 KFSAADGDLVVKASDIATHLNTLAGDIO-TAKGASQASSASYVDADGNKRVIVYDSTDPKY 1349  
Db 1270 ----adtda-atlkqvkeyrttlvgnditaaadrsqtsngitynlslnkgtvsateek- 1323  
QY 1350 YQVNDKQVDKNKEVAKDKLVAAQTPDGTTLAQMNVKSVINKEQVNDANKKQGINEDNAF 1409  
Db 1324 -----vvsqgktvyee-----frnait-----gnif 1343  
QY 1410 IKGLENAAKDTKTNAAVTVGDDLNVAQAQTPLTTFAGDTGTTA-----KKLGETLTIKG 1461  
Db 1344 tiglidd-----ttlnkinnpadqdlslslsesgknaitglvdvdkktnspitve- 1391

QY 1462 GQDTNKLTDNNIGVWAGTDCFTVKLAKD---LTNLSNVNAGGTRIDEKGISFYDANGQA 1518  
Db 1392 pstdsnkktftvg-vdftditegdatdckkittksve-----syv-tukla 1438  
QY 1519 KANTPVLISANGLDLGGKRISNIGAAVDNDNAVNFQKQNEVAKTVNNLN-----NQSNSGA 1573  
Db 1439 nfstdillisdg-----rsgnattandvgvgrlisd-gftiksenftlgsqyngsd 1488  
QY 1574 SLPEVVT'DANGKPI'NGT'DGK'PQKA'IKGADGK'YHANANGVPV'DKGK'PITDADKLANLAA 1633  
Db 1489 slgvmvdyddqng-----1499  
QY 1634 HGKPLDAGHGVVASLGSNSDAITLTINIKSTLPQIDTPTNTGNANAGQASLPSLSAAQOSN 1693  
Db 1500 -----vfklslnmtalt-tslantfakidasn-----ltdsdn 1531  
QY 1694 AASVKDVLNVGNFNLQTNHNOVDVKAYDVTNVFNGTGADITSRSADGTMSNTVNTVNTALA 1753  
Db 1532 kekwrtnlv-----ysktevdaeiqskvltltpdsglifatkgagsgnnagidagnkkl 1586  
QY 1754 ATDDDGHVLIKAKD---GKFYKADDLMPNGSLKAGKASDAKT--PTGLSLVNP- 1803  
Db 1587 snvadgdisptsgdvvtgrqlya--lmqgkiryvgydevspktqtaptanpattapt 1644  
QY 1804 -----AGKGSTGD-----AVALNLSKAVFK 1824  
Db 1645 asstqgwattantaggvapagnvatgdiaptqptlpemmtalvddhlaavlpggskli---1701  
QY 1825 SKDGTITTTTSSD---GISIQGKDNSSITLSKQGLNVGGKVISNVGKGTDKTDAANVOQL 1881  
Db 1702 hgdhnvkttisadnqvglisq--pnisie-----nnlvigsdnpekakiaa- 1745  
QY 1882 NEVRNLLGLGNAGNDNA-----DGNOVNIADIKKDPN---SGSSSNRTVIKACTVLGGK 1932  
Db 1746 -qegnalvtnkddgnaamfnnnekmlvlisdkeakprvlldgngaltiv-----1795  
QY 1933 GNNDTEKLATGGVQGVYDKDGNANGDLSNVVVKTKQDGSKALLATYNAAGQTYLTNNP 1992  
Db 1796 gnddsq--vtlsskkgkdidgn---dlslsvtttert-----nadgglekvetsf 1840  
QY 1993 AEADRINEOGIRFFHVNDGNEQPEVVOGRNIDSSASGKHSVAIGFOAKADGEAAVAGR 2052  
Db 1841 atmdd-----glkf-----kadgkvi-----1857  
QY 2053 QTOAGNOSIAIGDNAQATGQDSIAIGTGVNVVTGKHSIGAIGDPSTVKADNSYSVGNNOQFI 2112  
Db 1958 -nkklnetvel-----vgdenvt-----sitddnkvkvslnkklaiddevki 1898  
QY 2113 DATQDVFVGNNITVTESNSVALGSNSAISAGTHAGTQAKKSDGTAGTTTGTAGTGVK 2172  
Db 1899 pntdpda-qkgdsivi-----nngghagnkvitgvkasd-----1932  
QY 2173 GFAGQTAAGVAVSGASGAERRIQNVAAGEVSATSDTAVNSQLYKATQOGIANATNELDHR 2232  
Db 1933 -----dptsavnrqglnatvidnvnqnnfnqnr 1960  
QY 2233 IHQENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGOGAVAVGLSKLSDNGQWVF 2292  
Db 1961 igdltresagigamatasaqlnvalpgkttisvgtatfkgenavaigmsrlsdngkvgi 2020  
QY 2293 KINGSADTQGHVGAAGVAGGFHF 2314  
Db 2021 rlsgmstngdkgaamsvgfsf 2042

RESULT 15

AAB23858

ID AAB23858 standard; Protein; 1094 AA.

XX AAB23858;

XX 17-JAN-2001 (first entry)

DT

198 kvnvtstasglnfakgatgangdttvhlntastlqdtllntgvsvskldngnglcat-ekk 256  
 1201 RAASIGDVLNAGFNLOG-----NCEAVDFVSTYDVFIDGNATTAKVTYDDTS--KTS 1252  
 257 raasvqdvlnsgwnikgvtgattsdnvdvrtvdyotvflsgseettlvtvdsesngkst 316  
 1253 KVVYDVNDNKTEIVTSOKKLGKVTTLTKTSA---NGNATK-----FSAAD---GD 1298  
 317 kv-----kigaktsvikekgdklftgkankdtnqvnnaadtddegk 359  
 1299 ALVKASDIATHLNTLAGDIOTAKGASQA-----SSASVYVDADGN---KVIYDSTD-- 1346  
 360 glvtaetvlnavnkagwrkttgannqggfctvtsntvtfadngtgvttvdatngi 419  
 1347 --KYYQVNDKGOVDKNEKAVAKDLV-----AQATPDGCTLAQMNKVKSVINKEQVNDANK 1399  
 420 tvkyeakvvgdglikigndqkittadtaltvggkvtpadat---ngkklvnasgladeln 475  
 1400 KOGINEDNAFIKLENNAAK--DKYTKNAVTVGDLNNAVAQTPLTFAGDTGTTAKLGETLT 1458  
 476 k-----lswtakaeadtang---geid-----gtadekevkgagetvt 509  
 1459 IKGOTDTNKLTDNNIGVVAGTGDGTVKLAKDLTNLNSV-----NAGGTRIDEKGISPV 1512  
 510 fkgak-----nkvkqdganfityslqdalgtltsltlgtngngakteinkdgitit 560  
 1513 DANGQA--KANTPVLSANGLDGGKRISNIGAAVDDNDVAFKQFNEVAKTVNNLNQSN 1570  
 561 pangagannantistvdkingsaggsqkvvnvsgl-----kkfg----- 598  
 1571 SGASLPFVVTDANGKPIVGTGDKPQKAIKGADGKYIHANANGVVPVVDKGPITDADK-LA 1629  
 599 -----danfdptlss-----adnltkqydaykglt 624  
 1630 NLAARGPLDAGHGVAVSLOGNSDAITLTNIKSTLPDIDPNTGNAGAOQSLPSLSAA 1689  
 625 nldeg-----adkqtlvtadnt----- 642  
 1690 QOSNAASVKVNLVGNFNLQTNHNOVDFKAYDT-----VNFVNGTCAADITSVRSADGT 1742  
 643 ----aatvgdlrglvgwlsadktgeidkeynaqvrnanevkfksgnginv-sgktvng- 696  
 1743 MSNITVNTALAATDDGNVLKAKDG-----KFYKADDLMPNGSLKAGKSASDAK 1792  
 697 rreitfelakgevvksneftvketngketslvkvgdkyyskedldp-----atgkp----- 747  
 1793 TPTGLSLVNPNAKGSTGDAVALNLSKAVEKSKDGTGTTTTVSSDGIQOGKDNSSITLS 1852  
 748 -----kvngnava-----akyqdkg---kvvsadggs-----ntavlt 780  
 1853 KDGLN--VGGKVISNVGKGTKDTDAANVQQLNEVRNLLGLGNAGNDNADGNQVNTADIKK 1910  
 781 nkgygyvtgnqvadaiaak-----sgfeigladaek 810  
 1911 -----DPNSGSSNRTVFKAGTVLGGKGNNDTEKLATG---GVQGVVDKQGNNGD-LS 1960  
 811 akaafgetekalsd-----kletv---nandkvrfanglntkvsaaatesidangdkvt 862  
 1961 NVWVKTKDGSKALLATYNAAGQTNLYLTNNPAEAIDRINEQGIFFHVNDQNPVQVG 2020  
 863 ttfvktdeve---lplqtlyv----- 879  
 2021 RRGIDSSASGKHSVAIG---FOAKADGEAAVAIGRTQAGNQSIAIGDNAQATGQDSIAI 2077  
 880 -----tdangkkivkngdkwyyckdggst-----dmtkev 909  
 2078 GTGNVYT--GRHSIGAIGDPSTVKADNSYSVGNNNQFIDATOTDFVGVGNNTVTESNVAL 2136  
 910 tlgvndsgk-----kvvkednkwyhvksdgsdktq-----vveekvst 950  
 2137 GSNISAIAGTHAGTQAKKSDGTAGTTTGTAGATCTVKGFGAGTAVGAVSVGASGAEIRIQN 2196  
 951 dekhvvaldpndqsgk-----gvv-----inn 973

Db 198 kvnvtstasglnfakgatgangdttvhlntastlqdtllntgvsvskldngnglcat-ekk 256  
 QY 1201 RAASIGDVLNAGFNLOG-----NCEAVDFVSTYDVFIDGNATTAKVTYDDTS--KTS 1252  
 Db 257 raasvqdvlnsgwnikgvtgattsdnvdvrtvdyotvflsgseettlvtvdsesngkst 316  
 QY 1253 KVVYDVNDNKTEIVTSOKKLGKVTTLTKTSA---NGNATK-----FSAAD---GD 1298  
 Db 317 kv-----kigaktsvikekgdklftgkankdtnqvnnaadtddegk 359  
 QY 1299 ALVKASDIATHLNTLAGDIOTAKGASQA-----SSASVYVDADGN---KVIYDSTD-- 1346  
 Db 360 glvtaetvlnavnkagwrkttgannqggfctvtsntvtfadngtgvttvdatngi 419  
 QY 1347 --KYYQVNDKGOVDKNEKAVAKDLV-----AQATPDGCTLAQMNKVKSVINKEQVNDANK 1399  
 Db 420 tvkyeakvvgdglikigndqkittadtaltvggkvtpadat---ngkklvnasgladeln 475  
 QY 1400 KOGINEDNAFIKLENNAAK--DKYTKNAVTVGDLNNAVAQTPLTFAGDTGTTAKLGETLT 1458  
 Db 476 k-----lswtakaeadtang---geid-----gtadekevkgagetvt 509  
 QY 1459 IKGOTDTNKLTDNNIGVVAGTGDGTVKLAKDLTNLNSV-----NAGGTRIDEKGISPV 1512  
 Db 510 fkgak-----nkvkqdganfityslqdalgtltsltlgtngngakteinkdgitit 560  
 QY 1513 DANGQA--KANTPVLSANGLDGGKRISNIGAAVDDNDVAFKQFNEVAKTVNNLNQSN 1570  
 Db 561 pangagannantistvdkingsaggsqkvvnvsgl-----kkfg----- 598  
 QY 1571 SGASLPFVVTDANGKPIVGTGDKPQKAIKGADGKYIHANANGVVPVVDKGPITDADK-LA 1629  
 Db 599 -----danfdptlss-----adnltkqydaykglt 624  
 QY 1630 NLAARGPLDAGHGVAVSLOGNSDAITLTNIKSTLPDIDPNTGNAGAOQSLPSLSAA 1689  
 Db 625 nldeg-----adkqtlvtadnt----- 642  
 QY 1690 QOSNAASVKVNLVGNFNLQTNHNOVDFKAYDT-----VNFVNGTCAADITSVRSADGT 1742  
 Db 643 ----aatvgdlrglvgwlsadktgeidkeynaqvrnanevkfksgnginv-sgktvng- 696  
 QY 1743 MSNITVNTALAATDDGNVLKAKDG-----KFYKADDLMPNGSLKAGKSASDAK 1792  
 Db 697 rreitfelakgevvksneftvketngketslvkvgdkyyskedldp-----atgkp----- 747  
 QY 1793 TPTGLSLVNPNAKGSTGDAVALNLSKAVEKSKDGTGTTTTVSSDGIQOGKDNSSITLS 1852  
 Db 748 -----kvngnava-----akyqdkg---kvvsadggs-----ntavlt 780  
 QY 1853 KDGLN--VGGKVISNVGKGTKDTDAANVQQLNEVRNLLGLGNAGNDNADGNQVNTADIKK 1910  
 Db 781 nkgygyvtgnqvadaiaak-----sgfeigladaek 810  
 QY 1911 -----DPNSGSSNRTVFKAGTVLGGKGNNDTEKLATG---GVQGVVDKQGNNGD-LS 1960  
 Db 811 akaafgetekalsd-----kletv---nandkvrfanglntkvsaaatesidangdkvt 862  
 QY 1961 NVWVKTKDGSKALLATYNAAGQTNLYLTNNPAEAIDRINEQGIFFHVNDQNPVQVG 2020  
 Db 863 ttfvktdeve---lplqtlyv----- 879  
 QY 2021 RRGIDSSASGKHSVAIG---FOAKADGEAAVAIGRTQAGNQSIAIGDNAQATGQDSIAI 2077  
 Db 880 -----tdangkkivkngdkwyyckdggst-----dmtkev 909  
 QY 2078 GTGNVYT--GRHSIGAIGDPSTVKADNSYSVGNNNQFIDATOTDFVGVGNNTVTESNVAL 2136  
 Db 910 tlgvndsgk-----kvvkednkwyhvksdgsdktq-----vveekvst 950  
 QY 2137 GSNISAIAGTHAGTQAKKSDGTAGTTTGTAGATCTVKGFGAGTAVGAVSVGASGAEIRIQN 2196  
 Db 951 dekhvvaldpndqsgk-----gvv-----inn 973

XX DE Haemophilus influenzae adhesin (Hia) protein from NTHi strain K9.  
 XX KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;  
 KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;  
 KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;  
 KW diagnosis; immunogenic; antigen.  
 XX OS Haemophilus influenzae.  
 XX KW WO200055191-A2.  
 XX PN 21-SEP-2000.  
 XX PD 16-MAR-2000; 2000WO-CA00289.  
 XX PF 16-MAR-1999; 99US-0268347.  
 XX PR (CONN-) CONNAUGHT LAB LTD.  
 XX PA Loosmore SM, Yang Y, Klein MH;  
 XX PI WPI; 2000-618897/59.  
 XX PT N-PSDB; AAA92497.  
 XX DR Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for  
 XX use as antigens and vaccines and for treating Hemophilus influenzae  
 XX infection  
 XX Claim 1; Fig 22; 275pp; English.  
 XX The present sequence represents a Haemophilus influenzae adhesin (Hia)  
 CC protein from the non-typeable Haemophilus influenzae (NTHi) strain K9.  
 CC Hia genes and proteins have antiinflammatory, auditory and antibacterial  
 CC activities, and can be used in the production of a vaccine. An  
 CC immunogenic composition comprising an Hia gene, a polypeptide encoded  
 CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
 CC protection against disease caused by Haemophilus strains in a  
 CC susceptible host, preferably a human. An Hia protein is useful as an  
 CC antigen, in immunogenic preparations including vaccines, as a carrier  
 CC for other immunogens, and in the generation of diagnostic reagents. Hia  
 CC is useful for treating diseases caused by the infection of Haemophilus  
 CC influenzae such as meningitis, epiglottitis, septicemia and otitis  
 CC media. Recombinant production of Hia favours high recovery of the  
 CC protein compared to the low recovery of native protein from Haemophilus  
 CC influenzae species. A truncated protein has a significantly higher  
 CC amount of recovery than a full-length protein.  
 XX SQ Sequence 1094 AA;  
 Query Match 6.1%; Score 708; DB 21; Length 1094;  
 Best Local Similarity 23.0%; Pred. No. 1e-23;  
 Matches 345; Conservative 186; Mismatches 430; Indels 540; Gaps 67;  
 QY 915 TVNTLKLKENGTLVATNKDGTVTGINTQSLKAGDSTTLNKDGLSKNPASNEQIQVG 974  
 Db 33 tvatavlatqlsataeanssavsrlrvy-----gdCn-----kfaaansl--- 76  
 QY 975 ADGVKFAKVDKGNSSFGI-DGTSRTTKDQIGFTGANGSL--DTTKPHLTDKLKVGEVEI 1031  
 Db 77 -----adlnkqndgvhdgllninen-----gankkllvddntaatvgdlrklgwvvs 123  
 QY 1032 TNPGINAGGKKTNTIQSGD--ITQNSNDVATGGRVYDLKTELESKINSAAKTAONSLH-- 1087  
 Db 124 tkng--keneksqvkqadevifksgkv-----qvtstsengkhai 164  
 QY 1088 EFSVADEQGNHFTVPNPSYSDTSKTSDVITFAGENGIFTKVNKGVVRVGDQTKGLTTP 1147  
 Db 165 tfaIakd-----ldmrtatvsdltliggs--tt-----tgsattp 197  
 QY 1148 KLTV-----GNNNKGKIVIDSKDQWNTITGLSNTLAN-VTNDGAGHAL-SQGLANDTOKT 1200







Db 1 MNHIYKVFENKATGTFMAVAEYAKSHSHSGGSCATQGVGSRVTLSPFARIAALAVLIGAT 60  
 Qy 61 LNSGAYQONNSKIAPGTGNNNDNASASNEASTAIGSLAKAHANQAIAGGSKPDPNRQAA 120  
 Db 61 LNSGAYAGIGISEADGGK---GANARCDKSAIGADIAQALGSGSIAIGDNK-IVHNSN 116  
 Qy 121 NOKAGSHAKGRESIAIGSDVLAEGDASTAIGSDLLYLDNRNSTSKYPNGLLSTLIQHTV 180  
 Db 117 NANIKAKASGNEIAIGSDVLAEGDASTAIGSDLLYLLKKTVOO---ISELLPIIROOKA 173  
 Qy 181 LR---QIRDSNGSOKYRTUAEAGHASTAVGAMAYAKGHAFANAPGTRSTAEAGNYSLAVGLT 237  
 Db 174 LNDIYOLADTN-LQYRTHAGHASTAVGAMAYAKGHAFANAPGTRATAEGTYSILAVGLT 232  
 Qy 238 AKAEKYTTAIGSNAQAINYGALGADTRVDLDYALGYGSQL---NNNNNNNNKAY 294  
 Db 233 AYAKAASSIAGVSNQAIGFAATAVAGGSGTQVNLNRGIALGFGSOVLQKRDNDVNAANVRAY 292  
 Qy 295 VPEGNGSNKSKAT-GNG---LFGTSS---TIKRIINVGAGYEDTDVAVNVAQLKAV 346  
 Db 293 APDQNPIDNRYKATFKNGATDVFSIGNSGNDSTIRKLIINVGAGSADTDVAVNVAQLKEA 352  
 Qy 347 ENLAKQIITFKGDDNGTGVKKLGETLTITKGETQADKLTDNNNIGVYTDNNTGLKVKLA 406  
 Db 353 VRLANQIITFKGDDNSNRVEKGLGTLTITGG-AQTSALTD-HNIGVV-QNGDGLKVOLA 409  
 Qy 407 KNLGSETVSTKNLTASEKVTVGSGNNTAELQSGGLTFTPTTNASTVDKTVYGTGDLKFTD 466  
 Db 410 ETTLTSLKMWNTENLTANEKTVCK----- 433  
 Qy 467 NSNTALEDTTRITKDIGFSNKAAGTVDENKPYLDKD----- 502  
 Db 434 -----TRLTDLKIGTNDMNGIDESKPYLDKDTGTHAGGQKITKLTAAGVDDDAAT 484  
 Qy 503 -----KLK----- 505  
 Db 485 YGOLKKNQTAESALOTFTVKYKDKNGNDANDSKIIITVGNKNNKPGDTQVNTLKLKENGCV 544  
 Qy 506 -----VGNSTLNGGLTVNNTIGGSKNQIOVGADGKIFADVNVN 544  
 Db 545 DVTTETNGTTFGLNQNNGLTVGNSTLNDGLSVKNT---NSNQIOVGADGKITFTDISNS 602  
 Qy 545 VSNAAKFGTTRITEEIGFADAGKVDKSPYLDKQLQGVGVKTIYKDSGINAGDQKISN 604  
 Db 603 KPGAGIENTTRITRDGIGFANNTGSLDANKPRL-----TPGICNAGGKELTN 649  
 Qy 605 VKDATDOTDAVYTKLQKQVQDADGALQSPSIR-----DEKGOEFTISNLYNGHNTPN 658  
 Db 650 VQSA-----INPATNGGQLDFWNRLSTANTEKSGSAATIKDLYNLSQVP-- 693  
 Qy 659 FETITFAGENGISI-----SNDIAKGVKVGIDPT-NGLTTPKLVGSD 701  
 Db 694 ---LTFAGDTGPNVTKKLGEILKVGKGTADDLTKNIGVAVDSDNSLTVLAKTLS 750  
 Qy 702 KDG---KTQLVIEOVA---SGNDTKNI----- 722  
 Db 751 LDVAVNTKLTAADKTVDSGNNATAKLONGDLTFSKQNTGATPATNSKTIYGVGDLKFTDN 810  
 Qy 723 -----TRGLS 727  
 Db 811 NGIALDGTITYITKDKVGFAGQDGLSKSPYLDKDKLVGVEITTINGINAGGKAITGLS 870  
 Qy 728 PPLPSITNAGGVRTTEQONTITSDEKSKAASIGDILNTGFKLNKNSNSGFFVSTYNTVD 787  
 Db 871 NFLTDATNATTGHVTVQLG---IVDSTDKTRAASIGDVLNAGFLNKNNGDAKDFVSTYD 928  
 Qy 788 FIDGNATTAKVYDETNQTSKYTVDVNVDEKTIETLTDNGKINKIGYKTTTLTTTANG- 846  
 Db 929 FINGNATTAKVYD--GRASKVAYDVNVDTTILHTGADGNKNOIGVKTTLTTLTKDAKD 986  
 Qy 847 KATNFSTT---DNDALVNAKDAENLTLAKELHTTKGTADTALOTFKVKK-----DCATD 899  
 Db 987 KAINFVNSGDDKALINAKDIADNLTLAGEIRNTKGTADTALOTQFQVKVKENGDDND 1046

Qy 900 DETITVKGDTONGKTYNTLKLKGNGTLVATNKDGTVTFGINTQSGLKAGDSTLLNKDG 959  
 Db 1047 ADTITVKGDAKTN--QVNTLKLKGNGLDIQTNKDGTVTFGINTQSGLKAGNNTLLNNG 1104  
 Qy 960 LSTKNPASNEQIOVGADGVKFAKVDKGNSTGIDGTSRITKDQIGFTGANGSLDTTKPHL 1019  
 Db 1105 LSTKNTAGNEQIOVGADGVKFAKVNNGVYVAGIDGTTTRITRDEIGFAGTNGSLDKSKPHL 1164  
 Qy 1020 TRDKLVGVEITNTGINAGGKITINQSGDITQNSDAVTGRRVYDILKTELESINSAA 1079  
 Db 1165 SKD-----GINAGGKIIINQSGEIAQNSDAVTGKGIYDLKTELENKISSTA 1212  
 Qy 1080 KTAQNSLHEFSVADEOGNHFTVSNPYSSYDTSKTSVITFAGENGITTKVNGGVYRVGID 1139  
 Db 1213 KTAQNSLHEFSVADEOGNFTVSNPYSSYDTSKTSVITFAGENGITTKVNGGVYRVGID 1272  
 Qy 1140 QTKGLTTPKLTIVGNNGKGVIDSQDQNTITGLSNTLANVTND-GAGHALSQG-LANDT 1197  
 Db 1273 QTKGLTTPKLTIVGNNGKGVIDSQDQNTITGLSNTLANVTNDGKSVRTTEQGIIRDE 1332  
 Qy 1198 DKTRAASIGDVLNAGFNLOGEAVDFVSTYDIDGNATTAKVYDIDTAKVYDIDTAKVYD 1257  
 Db 1333 DKTRAASIVDVLNAGFNLOGEAVDFVSTYDIDGNATTAKVYDIDTAKVYDIDTAKVYD 1392  
 Qy 1258 VVNDKTIETVSDKKLGVTITLTKTSANGNATKF---SAADGDALVKASDIATHLNTLA 1314  
 Db 1393 VVNDKTIETVSDKKLGVTITLTKTSANGNATKF---SAADGDALVKASDIATHLNTLA 1449  
 Qy 1315 GDITAKGASQASASSYVDADGNKVIYDSTDKKYQVNDKQVNDKQVNDKQVNDKQVNDKQVND 1374  
 Db 1450 GDITAKGASQASASSYVDADGNKVIYDSTDKKYQVNDKQVNDKQVNDKQVNDKQVNDKQVND 1509  
 Qy 1375 TPDGTLAQMNVKSVINKEQVNDANKKOGINEDNAFKGLENAAKDTKKNAAVTVGDNA 1434  
 Db 1510 TPDGTLAQMNVKSVINKEQVNDANKKOGINEDNAFKGLENAAKDTKKNAAVTVGDNA 1569  
 Qy 1435 VAOTPLPAGDTGTAKKLGELTITKGGQDITKKNLTDNNIGVAGTDFGTVKLAKDLTNL 1494  
 Db 1570 VAOTPLPAGDTGTAKKLGELTITKGGQDITKKNLTDNNIGVAGTDFGTVKLAKDLTNL 1629  
 Qy 1495 NSVAGSTRIDKEGIFSDVANGQAKANTPYLSANGLDIGKRRISNIGAAVDDNDVAFKQ 1554  
 Db 1630 NSVAGSTRIDKEGIFSDVANGQAKANTPYLSANGLDIGKRRISNIGAAVDDNDVAFKQ 1667  
 Qy 1555 PNEVAKTVNNLNQNSGASLPFVVTDANGKPIGTGDKPKQALKGADGKYHANANGVP 1614  
 Db 1668 ----- 1667  
 Qy 1615 VDKGKPIITDADKLANLAHKGPLDAGHQVVASLGSGNSDAITLTNIKSTLPQIDTPNTGN 1674  
 Db 1668 ----- 1667  
 Qy 1675 ANAGQAQSLPSLSAAQSSNAASVKDVLNNGVFNLOTHNQVDFVKAYDTPVNFVNGTADIT 1734  
 Db 1668 ----- 1667  
 Qy 1735 SVRSADGTMSTNITVNTALAATDDDDGNVLTAKADGKFFYKADDLMPNGSLKAKSADAKTP 1794  
 Db 1668 ----- 1667  
 Qy 1795 TGLSLVNPNGKSTGDAVALNLSKAVPKSKDGTITTTVSSDGISIQGKDNSSITLSKD 1854  
 Db 1668 ----- 1667  
 Qy 1855 GLNVGKGVISNVKGTKDTDAANVQQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNS 1914  
 Db 1668 ----GKVLISNVKGTKDTDAANVQQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNS 1723  
 Qy 1915 GSSSNRTVIAKGTVVLGGKKNNDTEKLGATGGVQGVYDKNANCDLSNVVWTKQKDGSKKA 1974  
 Db 1724 GSSSNRTVIAKGTVVLGGKKNNDTEKLGATGGVQGVYDKNANCDLSNVVWTKQKDGSKKA 1783





1054 NSDAYTGRVYDVKTELESKINSAAKTAQNSLHEFSVADBOGNIHFTVSNPYSSYDTSKT 1113  
1223 -----LKYKDTQ-NTADETQKXPHA-----AVKN 1247  
1114 SDVITPAGBNGITTKV---NKGVRVIGIDQTK-----GLTTP-----KLJVGNNGKGIY 1160  
1248 ANEVFVGKNGATVSAKTDNNGKHVTITDVAEAKVGDGLERDGTGKIKLKYDNTDGNLL 1307  
1161 -IDSKDQNTITGLSN-----TLANVTN-DGAGHALSOG-----LANDTKTRAASICDV 1208  
1308 TVDATKGA5VAKGEFNAVTTDAATTAQGTNANERKGVVKGSGNGATATETDKKKVATVGDV 1367  
1209 LNAEFLNLOGNEAVFVSFYDTVDIFDGNATTAKYTYDDT-----SKTSKVYVDVYNVDNK 1263  
1368 AKA-----INDAATFVKVENDSDSATIDSPDGDGANDALKAGD 1405  
1264 TIEVTSKKLGVKVTITLTTSANGNATFESAADGDALVKA5DLATHLNTLAGDIQTAKGA 1323  
1406 TLTAKAGNLKVK-----RDGKNITFALAN-DLSVKSATVSKLS-----1444  
1324 SOASSASVYDADCNKVIYDSDKKYQVNDKGOVDKNEKAKDLVAQAQPPDGTLAQM 1383  
1445 -----LGTNGKNVITSDTKLNFPAKD-----SKTGGDANIHLNGIATLTDTLLNS 1491  
1384 NVKSVINKEOVNDANKQGINEDNAF-----IKGLENAKDKTKNAAVTVGDLNAVAQT 1438  
1492 GATTNLGGNGITDNEKKRAASVADVLNAGNVNKGVPASANNNOVENI-----DFVATYDT 1546  
1439 PLTFAGDTGTAKKLGTLTIKGGQDITKLNLTDDNITG-----VWAGPDG--FTVKLAKDLT 1492  
1547 VDFVSGDKDIT-----SVTVE--SKDNKRTVEKIGAKTSVIKDHNGKLTGKELKDOAN 1598  
1493 NLNSVNAAGTRIDKGTISFVDANGQAKANTPULSANGCLDL---GGRISNIGAAVDDNDA 1549  
1599 N-----NGVTVTETDCKDRGNGLVTAKAVIDAVNKAQWRVKTGTGANGONDD- 1644  
1550 VNFQFNEVAKTVNVLNNSQSGASLFPVYTDANGKPINGTDGKPOKA1KGA-DGKYIHA 1608  
1645 -----FATVAGTN-----VTEADG---NGTTAEVTKANDGSITVKYVNVK 1681  
1609 NANGVPVDKDKPITDADKLANLAHOK-----PLDAGHOVVASLGGNSDAITLTNIKSTLP 1665  
1682 VADGLKLGDG-KIVADTITVL--VADGKVTAPNNGDKKFEVDASGLADALNKLSWTAT-- 1736  
1666 QIDTPTGNANAGQAQSLPSLSAAQSNASVSDVLNAGFNLOTNHNOVD-----VK 1718  
1737 -----ACKEGTGEVD--PANSAGQEVKAGD-KVTERAGDNLKIKQSGKDFYSLAKKEL 1787  
1719 AYDTVNFVN---GTGADITSVRSADGMSNTTV---NTALAATDDGDNVLKAKDCKFYK 1772  
1788 DLTSVEFKDANGGTGSESTKI-----TKDGLTITPANGAGAAGANTANTISVTKDG----- 1838  
1773 ADDLMPNGLAKKASDASDAKPTGLSLVNPAGKSGTGDVALANLNSKAVFKS-----K 1826  
1839 -----ISAGNKA-----VTNVYVGLKKFGDGTLANGTVADFEKHVDNAYK 1879  
1827 DGTHTTTVSSDGISIQGKDNSSITLSKDLNAG-----GKVIS---NVGKTKDTDAAV 1878  
1880 DLT-----NLDEKGDADNPNTADNTAATVGDRLGLGWISADKTTGEPNQEYN-AQV 1930  
1879 QQLNEVRNLLGLG-NAGNDNAGNOVNIAIDKDP-----NSGSSSNRTVIKAGTV 1928  
1931 RNANEVFKFSGNGINVSQGLTNGTRVITFEKAKGEVKSNEFTVKNAQDSETNLVKVGDM 1990  
1929 LGKGKND--TEKLTAGGVQGVQVDKNGANGDLN5VWVKTKDGSKKALLATYNAACQTN 1986  
1991 YYSKEDIDPATSKPMTG-----KTEYKVKVNGKV-----VSANGSKTEVTTLTKSGS--- 2037  
1987 YLTNPAEADIRNEOGIRFEHFVNDGNQEPVVGQRNGIDSSASGKHSVAIGFOAK----- 2041  
2038 YVTGN--QVADAIKAGSGFEL-----GIADAAEAEKAEASAKDKQLSKD 2079  
2042 -----ADGEAAVAIGRQTAGNOSTAIGDNAQATGDQ5IAIGTGNVTVTKHSIGAGDPS 2095

2080 KAETVNAIDKVPFANGLTKV---SAATVESTDANGDK-----VTT-----T 2118  
2096 TVKADNSYSVGNNOFIDATQTDVFGV---GNNITVT-----ESNSVALGNSAISAG- 2145  
2119 FVKTD-----VELPLTQIYNTDANGNKIVKKADGWYELNADGTASKEVTLGN 2167  
2146 --THAGTOAKKSDGTAGTITTTAGATGTGKGFAGQTAAGVAVSGV-----ASGAE 2191  
2168 VVANGKVVVKTENGADKWWYTNADGAADTKGEVSDNKVSTDEKHVVRLDPNNQSNKG 2227  
2192 RRIQNVAAAGEVSATSDAVNGSOLYKATQGIAN---ATNELDHRHONENKANAGISSAM 2248  
2228 VVIDNVANGEISATSDAINGSQLYAVAKGVTNLAGQVNNLEKGVKVKRADAAGTASAL 2287  
2249 AMASMPQAYTPGRSMVTGGTATHNGQCAVAVGLSKLSDNGQWVFKINGSADTQGHVGA 2308  
2288 AASQLPQATPMGKSWAVAIAGSSVQGGNGLAIGVSRISDNGKVIIRLSGTTNSQKGTGVA 2347  
2309 GAGFHF 2314  
2348 GVGQVW 2353

RESULT 3  
US-08-913-942-4  
Sequence 4, Application US/08913942  
Patent No. 6200578  
GENERAL INFORMATION:  
APPLICANT: St. Gene, Joseph  
APPLICANT: Barenkamp, Stephen J.  
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESS: Flehr Hohbach Test Albritton & Herbert LLP  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,942  
FILING DATE: 29-DEC-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/409,995  
FILING DATE: 24-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/4031  
FILING DATE: 22-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Vance, Dolly A.  
REGISTRATION NUMBER: 39,054  
REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 398-3249  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2353 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-913-942-4



REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-61053/REF  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1912 amino acids  
TYPE: amino acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
US-08-409-995-4

Query Match 7.2%; Score 841; DB 1; Length 1912;  
Best Local Similarity 23.0%; Pred No. 1,6e-38;  
Matches 475; Conservative 255; Mismatches 657; Indels 676; Gaps 99;

QY 1773 ADDLMPNGLKAGKASDAKTTPTGLSLVNPAGKSGTGDVALNLSKAVFKS-----K 1826  
Db 1839 -----ISAGNKA-----VTNVVSGLLKFKDGHITLANGIVAFERKHYDNAYK 1879  
QY 1827 DGTITTTTSSDGSISQKDNSSITLSKDLNVG-----GKVIS---NVGKGTKDTDAANV 1878  
Db 1880 DLT-----NLDEKADNPTVADNTAATVCDLRLGLGWISADKTTGEPNOEYN-AOV 1930  
QY 1879 QQLNEVRNLLGIG-NAGNDNADGNQVNIADIKDP-----NSGSSNRRTVIKAGTV 1928  
Db 1931 RNANEVKEKSGNINVSCKTLNTRVITFELAKGEVVKNSNEFTVKNADGSETNLVKVCDM 1990  
QY 1929 LGGKGNND--TEKLTGGVGVGVGDGNGDGLSNVVKTKQDKGSKKALLATYNAAGOTN 1986  
Db 1991 YSKEDIDPATSKPWG-----KTEYKVENGV-----VSANGSKTEVTLTNKSGS--- 2037  
QY 1987 YLTNNPAEADIRNEOGIRFFHVNDGNQEPVVGQNGRIDSSASGKHSVAIGFOAK----- 2041  
Db 2038 YVTGN--QVADALAKSGFEL-----GLADAAAEKAFSAKDKQLSKD 2079  
QY 2042 -----ADGEAAVAIGROQAGNOSTAIGDNAQATGDQSIAGTGNVVTGKHSGAIGDPS 2095  
Db 2080 KAETVNAHDKVRPANGLNKTV---SAATVESTDANGDK-----VTT-----T 2118  
QY 2096 TVKADNSYSYGNNNQFIDATQTDVFGV---GNNITVT-----ESNSVALGNSAISAG- 2145  
Db 2119 FVKTD-----VELPLTOIYNTDANGNKKIVKADGKWYELNADGTASNKEVTLGN 2167  
QY 2146 --THAGTQAKSDGTAGTTTACATGVKGFAGOTAVGAVSVG-----ASGAE 2191  
Db 2168 VDANGKKVVKVTENGADKWTYTNADGAADTKGEVNDKVDSTDEKHVRLDPNNQSNKG 2227  
QY 2192 RRIQNAAGEVSATSTDAVNGSOLYKATOGIAN---ATNELDHRIHQENKANAGISSAM 2248  
Db 2228 VVIDNVANGEISATSTDAINGSOLYAVAKGVNTLAGOVNLEKGVKVKGRADAGTASAL 2287  
QY 2249 AMASMPQAYIGRSMVGTGATHTNGOGAVAGLSKLSDNQGWFEKINGSADTQCHVGAIV 2308  
Db 2288 AASQLPORTMFKSMVAIAGSVOYQGGGLAIGVSRISDNKGKVIIRLSGTTNSQKGTGVA 2347  
QY 2309 GAGPHF 2314  
Db 2348 GVGQW 2353

RESULT 4  
US-08-409-995-4  
Sequence 4, Application US/08409995  
Patent No. 5646259  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen I.  
APPLICANT: St. Geme III, Joseph W.  
TITLE OF INVENTION: Haemophilus Adhesion Proteins  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/409,995  
FILING DATE: 24-MAR-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.

1 MNHIYKVFIPKATGTFMAVAEACAKSHSG-----SSSSTAGQVGVSSPV 43  
Db 1 MNKIFNVVNVMTQTWVVVSELRTHTKRLNRNGDPVLATLLFATVQANATDEDELDVP 60  
QY 44 IRLTRVATL-----AILVIGA-----TL 61  
Db 61 VRTAPVLSFSDKEGTGEKEVTENSNGIYFDKNKGVLKAGATLTKAGDNLKXKQXTDEX 120  
QY 62 NSGAYON-----NSKIAFGTTGNN-DNASASNEASIAIGSLAKAHANQIAI 108  
Db 121 NASFTYSLKKDLTDLTSVATEKLSFGCDKVDITSDANGLKAKTNGNVHLN---GL 177  
QY 109 GSKPD-----PRQANOK-----AGSHAKGESIAIGGOVLAEGDA 146  
Db 178 DSTLPDAVNTGVLSSTPFPNOVEKTRAAVTKVDLNGWNIKAKT---AGNV----- 229  
QY 147 SIAIGSDDLVLDNRSTNSKYPNGLLSTLIQNHVTLROI RDSNGSQKYRRTAAEGHASTAV 206  
Db 230 -----ESVDLVSAYN--NVEFTIGDKNTL---DVLTAKENXKTEVKFPTKSVI 279  
QY 207 GAMAYAKGHFANAGTRTAEGNYSVAVGL-TAKA-----EKGYTI-AIGSNAOAINYG 258  
Db 280 GKLFTEKGNNDTNKVTSTATDNTDEGNGLVTAKAVDANKAGWRVKTATTANGCNDPA 339  
QY 259 ALALGADTRVLDYDIALGYGSGQILNNNNN--NNKAYVPEGNGSNKSSK----- 307  
Db 340 TVASG--INVTFESGD--GTTASVTKDTNGNGITVKYDAKVGDLKFDSDKRIYADTTAL 395  
QY 308 -ATGNLFSIGSSTTKRLIINV-----AGYEDTDAVNVVAQLKAVENLAKRQ 353  
Db 396 TVTGGKVAEIAKEDDKKKLVNAGDLVTALGNLSWKAKAEADTDG-----ALEGIS 448  
QY 354 ITFKGDDNGTGVKKLIGETLTIKGGE-----TQADKLTDNNNIGVVTDDNTG 400  
Db 449 -----EVRKAGEITVFKACKNLKVKODGANFTVSLQDALTGITSLGCTTNGG 496  
QY 401 LKVKLAKNLSGLE-----TVSTKNLTASEKVTVGSCNNTAELOQSGL-----TFTP 446  
Db 497 NDAKTVINKDGLTITPAGNGGTGTNTISVT-KDGLKAGNKAITVWASGLRAYDDANFDV 555  
QY 447 TTNASTDKTVYGTGDLKFTDNSN-----TALEDTRITKTD---KIGF--SNKAGTVDE 494  
Db 556 LNNASDTLNRHVEDAYKGLLNLEKNANKQPLVTDSTAATVGLRLKLGWVSTKNGTKEE 615  
QY 495 NKPYLDKDKL---KVGNSLT-----NNGGLTVNNTIGGSNKQIOVGNDG--IKFADVNVV 545  
Db 616 SNQVKQADEVLETFGAGAAVTVTSKSENGKHTITVSVAEKTRKADCGLEKDGOTIKLKVNDQNT 675  
QY 546 SNAAKECT--TRITE-----EEIGFADAD-GKVDKSPYLDKKQLQVGVKTKKSGINA 597  
Db 676 DNVLTVGNNGTAVTKGGFEIVTKGTATDADRKV-----TVKDATAND 717  
QY 598 GDQKISNVKD-ATYDDTDAVTVYKQLKQVQDADGALQSFSIRDEKGOEFTISLNSYNG--N 654

Db 718 ADKKVATVKDVATVAINSAATFVKTEMLTTSIDE-----DNPTDNGKDD 760  
QY 655 TPNTFETITF-AGEN-----GISISNDIAG-KVKV-----GIDPINGLT-TPK 695  
Db 761 ALKAGDTLTFKAGKLNKVKRDGNITFDLAKNLEVKTAQVSDTLTIGGNPTGTATPK 820  
QY 696 LTVGSDDKQTLVIOVASGNDTKNI-IRGLSPTL--PSITNAGGVRTTEQONTITSDE 752  
Db 821 VNITSADGLN--FAKETADASGKNVYLKGIATTLTPE-----AGAKSSHVDLNDATK 874  
QY 753 DSKAASIGDILNTGNFKNNNSGVFSYNTVDFIDGNATTAKVTYDNTOTSKVTYD 812  
Db 875 -KSAASIEDVLRACNWIQNGNNVDYVATYDVNFTDDSTGTTTVV--TQADCKGAD 931  
QY 813 VNYDEKTIETGDKGNKTKIGVKTTLTTTNNANGKATNFSTTDDND-----ALVNAKDI 865  
Db 932 VKIGAKTSVTKDNGK-----LFTGDKLDKDNANGATVSEDDGKOTGTGLVTAKT 981  
QY 866 AENLNTLAKEIHTTKGTADTALQTFVKKDKGAT-----DDETITVKGDTGONTKVTNLT 920  
Db 982 IDAVNKGSRVTVGEGATAET-----GATAVNAGNAETVT-----SGTSVN-- 1021  
QY 921 LKGENGLTVATNKG--TVTFGINTOSGLKAG-----DSTLLNKDGLSIKSPA-- 966  
Db 1022 FKNGNATTATVSKONGNINVKYDVNVCDGLKIGDKKIIVADTTTLTVTGKVSVPAGANS 1081  
QY 967 -SNEQIQVGADGVK-----PAKVDK----- 985  
Db 1082 VNNKKLVNAEGLATALNLSWTAKADKYADGESEGTDOEVKAGDKVTPKAGKLNKVKQ 1141  
QY 986 -----GNSSTGIDTSR-----IYKDIQIGFTGANGSLDTPPHLTCKD 1023  
Db 1142 SEKDFYSLQDTLTGTSITSLGTGANGRNDGTGVINKDGLTITLANGAAAGT-----DA 1195  
QY 1024 LKVGVEITVTGNAGKKTNIQS-----GDIPTQ-----NSNDA-- 1058  
Db 1196 SNGNTSVTKDGLSAGNKEITNVKSALKYTKONTADETQDFEFAAANKANEVEFVGK 1255  
QY 1059 -----VTGGRVYD-LKTELESKIN-----SAAKTAQ 1083  
Db 1256 NGATVSAKTDNNGKHTVTIDVAEAKVGDLGKDTGDKLKLVKVDNTDGNLLTVDATKGAS 1315  
QY 1084 NSLHFSV-----ADQGNHFTV-SNPYSYDTSK-----TSDVIT 1118  
Db 1316 VAKGEFNAVTTDATTAGTNNANERGVVVKGSNGATATETDKKKVATGVGVAKAINDAAT 1375  
QY 1119 FA-----GENG-----ITTKVNGK--VVRVGDIDQTKL----- 1144  
Db 1376 FVKVENDDSATIDSDPDGDANDALKAXDTLTLKAGKLNKVKRDKGNITFALANDLSVKS 1435  
QY 1145 --TTPKLTGVNNGK-----GIVDSKDGQNT---ITGLSNTLAN-VTNDGAGHA 1188  
Db 1436 ATVSDKLSLGTNGKNVITSDTKGLNFAKDSKTGDDANIHLNGIASTLTDLLNSGATTN 1495  
QY 1189 LSOGLANDTUKTAAASIGDVLNAGFNLOG-----NGEAVDFSVTYDVFIDGNATT 1240  
Db 1496 LGNGITDNEKKAASVKVDLNAWGNVRGVKPAANNQVENIDFVATYDVFVSGDKDT 1555  
QY 1241 AKVTYD--DTSKTSKVVYDVNVNDNKTIEVTSDDKGLGVKTTTLTK-----TSANG 1287  
Db 1556 TSVTVESKONGKTEV-----KIGAKTSVVKDNGKLTGTGKELKDANN 1598  
QY 1288 NATKFAAD-----GDALVRASDIATHLNTLAGDIOTAKGASQ-----ASSSASYVDADG 1337  
Db 1599 NGVTVTETDCKDEGNLVTAKAVIDAVNAGWRVKTGTGANGQNDPFATVASGTVNVTADG 1658  
QY 1338 NKVIYDSTDKKYYQVNDKG--QVDKNKEVAK-----DKLVAQA--QTPDGTLAQNNVKS 1387  
Db 1659 N-----GTTAEVTKANDGSTVTKYNTKVAADGLKLDGDKIVADTVTLTVADGKVTAPN-- 1710  
QY 1388 VINKEQVNDANKOGINEONAFKIGLENAKDTKTKNAAVTVGDLNNAVQOTPLTEAGDTG 1447  
Db 1711 -----NGXGKK---FXDASGLAGCLNKLXSTATACKECT-CEVD-----PANSAGQ-- 1752

QY 1448 TTAKKILGETLTIKGGQDTNKLTDNNIGVVAGTGDGFTVKLAKDLTNLSV-----NAG-- 1500  
Db 1753 --EVKAGDKVTFRAG-----DNLIKIQSKXKDFYSLKKELKDLTSVEFKDANGGTG 1801  
QY 1501 --CTRDEKGISFVDANGQA-----KANTPVLSANGLDLGGKRSNI-----GAAV 1544  
Db 1802 SESTKTKDKGLTITPANGAGAAGANTANTISVTDKGISAGNKAVTNVYSLKKFKFGDGHLL 1861  
QY 1545 DDNDVNF-KOFNEVAKTVNNLN 1566  
Db 1862 ANGTVADFEKHYDNAYXDLTNLD 1884

## RESULT 5

US-08-685-467-4  
; Sequence 4, Application US/08685467  
; Patent No. 6060059  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme III, Joseph W.  
; APPLICANT: Barenkamp, Stephen J.  
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/685,467  
; FILING DATE: 22-JUL-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/409,995  
; FILING DATE: 24-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-61053-2/RET/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1912 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-685-467-4

Query Match 7.2%; Score 841; DB 3; Length 1912;  
Best Local Similarity 23.0%; Pred. No. 1.6e-38;  
Matches 475; Conservative 255; Mismatches 657; Indels 676; Gaps 99;

QY 1 MNHIYKVIENKATGTMAVAECAKSHSGG-----SSSSTAGVGSSPV 43  
Db 1 MNKIFNVIMVMTQTWVWVSELTRTHKRLNRNGDPVLATLLFATVQANATDEDELDPV 60  
QY 44 IRLTRVATL-----AIIIVGA-----TL 61  
Db 61 VRTAPVLSFHSDEKTEGTEKEVTENSNWGIYFDNKGVLKAGAITLKAGDNLKXKQTXDXT 120  
QY 62 NGSAYAQN-----NSKIAFGTTGNN-DNASASNEASIAIGSLAKAHANQAIAI 108





MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685.467  
FILING DATE: 22-JUL-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/409,995  
FILING DATE: 24-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robln M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-61053-2/RET/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1098 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-685-467-2

Query Match 5.68; Score 659.5; DB 3; Length 1098;  
Best Local Similarity 22.9%; Pred. No. 8.5e-29;  
Matches 323; Conservative 168; Mismatches 457; Indels 461; Gaps 55;

QY 1008 ANGSLDTKPHLTQKDKLV---GEVEINTGINAGGKKTINIQSDDITON----- 1054  
DB 49 ANNPVPT-----NKLKAYGANFNFNNSIADAEKQVQYKGLLNLEKNASDKLLV 102  
QY 1055 -SNDVATGGRVYDLKTELESKINSAAKTAQNSLHEFSVADEQNHFTVSNPYSSYDTSKT 1113  
DB 103 EDNTAATVGNLRKLGWLSSKNGTRNEKSSQVKKHAEVLFEKGKGVQVT---STSENGKH 159  
QY 1114 SDVITFAGENGITTKVNGVVRGVDQTKG-LTTPKLV-GNNCKGVIDS--KDGQNT 1169  
DB 160 TITFALAKDLGVKATVSTDTLIGGGAAGATTTPKVNVTSTDTGLKFAKDAAGAGDGT 219  
QY 1170 I--TGLSNTLANVNDGAGHALSOGLANDTDKTRAASIGDVLNAGFNLO-----GNG 1219  
DB 220 VHLNGIGSLTD-TLVGSPATHIDGGDQSTHYTRAASIKDVLNAGWNKGVKAGSTTGOS 278  
QY 1220 EAVDFVSYDTVDFFDGNATTAKYVYDDTSKTSKVYDVVNDKNKTIETVSOKKLGVKTTT 1279  
DB 279 ENVDFVHTYDVFELSADEFTTFTVTVDSKENGKRTVEVKIGAKTSVIKEDGKLFCKRANK 338  
QY 1280 LTKTSANGNATKFSAADGDALYKASDI-----ATHLNTLAGDIQT-AKASQA 1326  
DB 339 ETNKVDGANATE-DADEGKGLVATKDVIDAVNKTGWRIKTTDANGQNGDFATVASGTWVT 397  
QY 1327 SSSASYDA-----DGNKVYDYDTSKKYQVNDKQVQDNKKEVAKDLVAQAQPTDGTFL 1380  
DB 398 FASGNGTTATVTNGTGTGVKYDA-----KVGDLGLKD-----GDKIAA-----DTTA 440  
QY 1381 AQMNKSVINKQVNDANKKQGINEDNAFKGLENAADTK---TKNAAVTVGDILNVAQ 1437  
DB 441 LTVN-----DGNANNPKGVADVASTDEKRLVAKGLVTA--LNSLSW 482  
QY 1438 TPLTFAGDTGT-----TAKKLGETLITKGGOTDNTKLTDDNIGVWAGTDGFTVKLAKD 1490  
DB 483 TTTAAEADGGLDGNASEQEVKAGKDVFKAGK-----NLVKQEGANFTYSLQDA 533  
QY 1491 LFNLSV-----NAGGTRIDPKGISFVDANGQA--KANTPVLSANGLDLOGKRSINICA 1542  
DB 534 LTGLTSITLGTGNGAKTEINKDGLTITPANGAGANNATISVTKDGISAGSQSVKNVVS 593

QY 1543 AVDDNDVAVNFKQFNEVAKTVNNLNQNSGASLSPFVVTDANGKPIINGTDGKPKQAIKGDAD 1602  
DB 594 GLKKEGDAN---FDPLTSSADNLTKQND-----AYKG-- 623  
QY 1603 KYYIHANANGVPVDKGRKPTTDADKLANLAHCKPLDAGHQVVASLGGNSDAITLITNKS 1662  
DB 624 -----LTNLDE-----KGTDKQTPVYAD----- 641  
QY 1663 TLPQIDTPTNTGNANQAQSLPSLSAAQOSNAASVKDVLNVGNFLQTNHNVDFEYKAVDT 1722  
DB 642 -----NTA-ATVGDRLGLGWISADKTTGGSYE-----YHQ-----VRNAME 678  
QY 1723 VNFYNGTGADITSRSADGTSNITVTALAAATDDGNNVLIKAKDG-----KPYK 1772  
DB 679 VKFKSGNGINV-SGKTVMG-RREITFELAKGEVWKSNEFTVKTNGKETSLVKVGDKVYS 736  
QY 1773 ADDLMPNGSLKAGKASDAKPTGLSLVNPNAKKGSTGDAVALNNLSKAVFPKSKDGTITT 1832  
DB 737 KEDI-----DLTGQPKLKGNTVA 756  
QY 1833 TVSSDGIStQCKDNSSITLSKDLNVGKVISNVGKGTKDTDAANVQOLNEVRNLLGLCN 1892  
DB 757 AKYQD-----KGCKVVS---VTDNTEAT-----ITN 779  
QY 1893 AGNDNADGNQVNIADIKKDPNCGSSNRTVIKAGTVLGKGGNNDTEKLTATGQVGVVDKD 1952  
DB 780 KGSYVVTGNQV-----ADAIKASGFELGLADE 806  
QY 1953 GNANGDLSNVVVKTKDGSKKALLATYNAAGOTNYLTNNPAEADIRINFGIRFFHVNDG 2012  
DB 807 ADA-----KRAFDKTKAL-----SAGTTEIV-----NAHDK-----VRF----- 836  
QY 2013 NOEPVVGRRNGIDSSASGKHSVAIGFQ-----AKADGEAAVAIGRTOAGNOSI--AIGD 2065  
DB 837 -----ANGLNTKVSAAVESTDANGDKVTTTFTKTDVDELPLTYNTDANGKIKTVVKD 891  
QY 2066 -----NAQATGQOSTAIGTGNVVTGKHSIGAIGDPSTVKADNSYSVGNNOFIDATQT 2117  
DB 892 GOKWYELNADTADMTKEVTILGNVS-----DGKVVVKDND----- 928  
QY 2118 DVFGVGNITVTESNSVALGNSALSAGTHAGTQAKKSDGTAGTTTATAGCTVKGFAQO 2177  
DB 929 -----GKWHVHA-----KADGTADKT-----KGEVSN 949  
QY 2178 TAVG-----AVSVG-----ASGAERRIQNVAAGEVATSTDAVNGSOLYKATOGIAN---A 2225  
DB 950 DKVSTDEKHHVSLDPNDQSKGVVIDNVANGDISATSTDAINGSOLYAVAKGVTNLAGO 1009  
QY 2226 TNELDRHITHONENKANAGISSAMAMASHMPOAYIPGSRVMTGGIATHNGGAVAGLSKLS 2285  
DB 1010 VNNLECKVNVKGRADAGTASALASQLPQATMPGKSMVATAGSSYQONGIAIGVSRIS 1069  
QY 2286 DNGQWVFVINGSADTQGHVGAAGVAGFHF 2314  
DB 1070 DNGKVIIRLSGTTNSQGTGYAANGVGYOW 1098

RESULT 8  
US-09-377-155-32  
; Sequence 32, Application US/09377155  
; Patent No. 6197312  
; GENERAL INFORMATION:  
; APPLICANT: PEAK, Ian Richard Anselm  
; APPLICANT: JENNINGS, Michael Paul  
; APPLICANT: MOXON, E. Richard  
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN  
; FILE REFERENCE: 065064/0128  
; CURRENT APPLICATION NUMBER: US/09/377,155  
; PRIOR FILING DATE: 1999-08-19  
; PRIOR APPLICATION NUMBER: PCT/AU98/01031  
; PRIOR FILING DATE: 1998-12-14  
; PRIOR APPLICATION NUMBER: GB 9726398.2  
; PRIOR FILING DATE: 1997-12-12



QY	1833	TVSSGSIQKDNSSITLSKGLNVGVGKIVNVGKGTGKTDAAVNVQQLNEVRLGLGN	1899
DB	757	AKYQD-----KGGKWS-----VTDNTEAT-----ITN	779
QY	1893	AGNDNADGNQVNIADIKKDPNSGSSSRNTRVIRAGTVLGGKGNNDTEKLATGGVQVGVGDK	1952
DB	780	KSGVVTGNV-----ADAIKSGFELGLADE	806
QY	1953	GNANGDLNVVYKTKDGSKKALLATYNAAAGOTNVLTNPAEAIQRINEQIGRFFHNDG	2012
DB	807	ADA-----KRAFDDKTKAL-----SAGTTEIV-----NAHDK-----VRF-----	836
QY	2013	NOEPVVQVRNGIDSSASGKHSVAIGFQ-----AKADGEAAVAIGRTQAGNOSI---AIGD	2065
DB	837	-----ANGINTKVSAASTVESTDANGDKVTTTFVKTDVLPLOIYNTDANGKKITKWKD	891
QY	2066	-----NAQTGDSIAITGNNVTVGKHSIGAIDPSPVTKADNSVSVGNNOQFIDATQT	2117
DB	892	GQTKWYELNADGTADMTKEVTLGNVDS-----DGKKVVDND-----	928
QY	2118	DVFGVGNITVTESVALGNSAISAGTHAGTQAKKSDGTAGTTTAGATGTVKGFAG	2177
DB	929	-----GKWYHA-----KADGTADKT-----KGEVSN	949
QY	2178	TAIVG-----AVSVG-----ASGAERRIQNVAAGEVSATSDAVNGSOLYKATOGIAN---A	2225
DB	950	DKVSTDEKHVSLDPNDQSKGKGVVDNVANGDISATSDAINGSOLYAVAKGVTNLAG	1009
QY	2226	TNELDHRHONENKANAGISSAMAMAPQVYIPGRSWVTGGIATHINGOGAVANGLSKLS	2285
DB	1010	VNNLEGVKNVGRADAGTASALAASQLPQATMPKSMVAIAGSSYQOGQGLAIGVSRIS	1069
QY	2286	DNGQWVFKINGSADTQGHVGAAGVAGFHF	2314
DB	1070	DNGKVIIRLSGTTNSQGGTGVAGVGYQW	1098
<p>           RESULT 9            US-08-913-942-2            ; Sequence 2, Application US/08913942            ; Patent No. 6200578            ; GENERAL INFORMATION:            ; APPLICANT: St. Geme, Joseph            ; APPLICANT: Barenkamp, Stephen J.            ; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS            ; NUMBER OF SEQUENCES: 19            ; CORRESPONDENCE ADDRESS:            ; ADDRESSEE: Flehr Hohbach Test Albritton &amp; Herbert LLP            ; STREET: Four Embarcadero Center, Suite 3400            ; CITY: San Francisco            ; STATE: California            ; COUNTRY: United States            ; ZIP: 94111-4187            ; COMPUTER READABLE FORM:            ; MEDIUM TYPE: Floppy disk            ; COMPUTER: IBM PC compatible            ; OPERATING SYSTEM: PC-DOS/MS-DOS            ; SOFTWARE: PatentIn Release #1.0, Version #1.30            ; CURRENT APPLICATION DATA:            ; APPLICATION NUMBER: US/08/913,942            ; FILING DATE: 29-DEC-1997            ; CLASSIFICATION: 514            ; PRIOR APPLICATION DATA:            ; APPLICATION NUMBER: US 08/409,995            ; FILING DATE: 24-MAR-1995            ; PRIOR APPLICATION DATA:            ; APPLICATION NUMBER: PCT/US96/4031            ; FILING DATE: 22-MAR-1996            ; ATTORNEY/AGENT INFORMATION:            ; NAME: Vance, Dolly A.            ; REGISTRATION NUMBER: 39,054            ; REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV            ; TELECOMMUNICATION INFORMATION:         </p>			



```

; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 679 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-913-942-15

Query Match 5.1%; Score 592; DB 4; Length 679;
Best Local Similarity 23.9%; Pred. No. 2.4e-25;
Matches 221; Conservative 121; Mismatches 285; Indels 298; Gaps 32;

QY 1433 NAAQTPLETFAGDTGTTAKLGTLL-----TIKGGQTDTKLTDNNIGVAGTD 1481
Db 10 NVVTQWVVSSELTRHTKCSATVAVAVLATLLSATVQANATDENE----- 56
QY 1482 GFTVKLAKDLTNLSVNAAGTRIDEKISFVDANGAKANTPVLSANGLDLGGKRISNIG 1541
Db 57 -----DDEELEPVQSVLRWSFK--SAKEGTGEQEGTTEVINLN-TDSSGNAVGSST 106
QY 1542 ANVDDNDVNFQFN-----EVAKTYNNLNQNSGASLFFVYTDANGKINGTDGPKQK 1596
Db 107 ITFKAGDLNLIKQSGNDFYTSKELKNLTSVETELSF-----GANGNKVDIT- 155
QY 1597 AIKGADKYYHANANGVPVDKGPITDADKLANLAHKGPLDAGHQVVASLGNSDAIT 1656
Db 156 -----SDANGLKLAKTNG-----NGQNSNVH 177
QY 1657 LTNKTSPLQIDPTNTGNANAGQSLPSLSAAQSSNAASVKDVLNVGNFLQTNHNOVDF 1716
Db 178 LNGIASTL-----TDLAGTGTGHVDT--NIDAVNYHRAASVQDVLNSGWNIOGNNVDF 231
QY 1717 VKAYDTVNFVNGCADITSVRSADGTSNITVNTALAA-----TDDGNVLIKAKDGKFY 1771
Db 232 VRTYDVTDFVNGANANVSV--TADTAHKTTVRVDVTGLPVQVYVTEGKTVVKV-GNEY 288
QY 1772 KADDLMPNGSLKACKSASDAKTPGLSLVNPNAKGSTGDAVALNLSKAVEKSKDGT 1831
Db 289 KARD-----DGSADMNQKVENGELAKTKVKLVSASGT 320
QY 1832 TTVSSDGIISQKDNSSITLSKDLNNGVGGVINSVNGKGTDKTDAAVQQLNEVRNLLGLG 1891
Db 321 NPVK-----ISNVADGTTEDTDAVSPKQLKALQD----- 348
QY 1892 NAGNDNADGNQVNIADIKDPNCGSSNRVRIRAGTVLGGKGNNDTEKLTATGGVQGVGDK 1951
Db 349 -----KQVTL-----STSN-----AYANG----- 363
QY 1952 DGNANGDLSNWKTKDQKSKALLATYNAAGOTNYLTNNPAEIDRINEQGIREF-PHVN 2010
Db 364 -----TDNDGK-----ATQTLN-----GLNFKFKSS 386
QY 2011 DGNQEPVVOGRNGIDSSASG-----KHSVAIGFOAKADGEAAVAIGRQTQAGNQSIA 2062
Db 387 DGEL-----LKISATGDTVTFTPKGVSQVQ-----DDGRASISKGANTTEG--LVE 431
QY 2063 IGDNAQATGDSQTAIGTGNVVTGKHSA-----IGDPTVVRADNSYSV---GNNNQFI 2112
Db 432 ASELVESLNKLGKMKVGVKGVSGELDGTSKETLVKSGDKVTLLKAGDNLKVKQEGTN--FT 489
QY 2113 DATQDVFEGVGNNTVTFESNSVALGSNSAISAGTHAGTQAKKSDGTAGTTTGTAGATCTVK 2172
Db 490 YALKDELTVGKS-----VEFKDTANGANGASTKTKKDLGTLTILANGANGATVTT----- 537
```

```

QY 2173 GFAGQTAVGAVSVGASGAERRIONVAAGEVSATSTDAVNGSOLYKATQGIAN---ATNEL 2229
Db 538 ---DADKIKVAGDISAGNKAVKNVAAGEISATSTDAINGSQLYAVAKGVTNLAGQVNNL 594
QY 2230 DHRHONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNCOGAVAVCLSKLSDNQ 2289
Db 595 EGVNKKVGRADAGTASALAASQLPQATMPGKSMVSTAGSSYQOGNGLAIGVSRISDNKG 654
QY 2290 WVEFKINGSADTQGHVGAAGVAGGFHF 2314
Db 655 VIIRLSGTTNSQKGTGVAAGVGQW 679

RESULT 11
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-10

Query Match 4.8%; Score 565.5; DB 2; Length 1600;
Best Local Similarity 22.1%; Pred. No. 2.3e-23;
Matches 413; Conservative 242; Mismatches 709; Indels 505; Gaps 90;

QY 1 MNHIYKVFNKATGTFMAVECAKSHSGSSSTAGOVSSPVRLTRVATLAILVTGAT 60
Db 1 MNRIIRLFKSKRLNALVAVSELTR---GCDHST--EKGSEKPVV-TKVRHLALKPLSAI 53
QY 61 LINGSYAAQ-NNSKIAFTCTTGNNDNASNEASIAIGSLAKAHANQAIAIGSKPDPNQA 119
Db 54 LLSLGMASIPQSVLASLQG-----MSVVHGTATMQVDGNKTTIRN-- 94
```

QY 120 ANQKAGSHAKGESIAIGGDVLAEGDASTAIGSDDLYLDRN-----STNSKYPNGLL 171  
Db 95 -----SVNAIINWKFQIDQEMEQFLOESSNSAVENRVT 129  
QY 172 SYLIQHIVLROIRDSNGSOKYRPTAAEGHASTAVGAMAYAKGHAFANAFG-TRSTAE-GN 229  
Db 130 SDQISQ---LKGILDSNG-QVELINP-----NGITIGKDAIINTNGEFTASTDLISN 176  
QY 230 YSL-AVGLTAKAEKGVYTAIGSNAQAINVAGALAGADRVLDLYG-----I 274  
Db 177 ENIKARNFLEOTKDKAL-----AEIVNHGLITVKGDSVNLIGGVKNEGVISVNGSSI 231  
QY 275 ALGYSQIILNNNNNNNNKAY-----VPEGNSNITKSKATGNGLFSGSSTIKRK----- 324  
Db 232 SLLAGQKITISDIINPTITYSIAAPENEAINEALGDIFAKG-GNINVRAATIRNKKLSADS 290  
QY 325 -----IINVAGYEDTDVNVNOLKAVENLAKRQITFKGDDNGTGVKKLGETLTI 375  
Db 291 VSKDSGNIVLSAKEGEAEIGGVISAQNOQAKG-GKLMIT-----GOKVTLTKTGAVIDL 343  
QY 376 ---KGETQ--ADKLTDDNNIGVWTDNNTGLVKLAKNLISGLETVSTKNLTASER----- 425  
Db 344 SKEGGETYLGDERGEGN-----GLOLAKTT-LEKGSTINVSKEKGGRAI 391  
QY 426 ---VTVGSGNNTAE-----LQSGGLTFTPTTNASTDKTVYGDGLKFTDWSNTALEDDT- 476  
Db 392 VWGDIALIDGNINAQGSDIARTKGFVETSGHDSIGDDVIDDAKEWLLDDPDVSIETLTS 451  
QY 477 --RITKDKTGFNSKAGTVDE-----NKPYLKDKLK-----VGNSTLNNGGTIVNNT 521  
Db 452 GRNNTGENGYITGDTKESPKGNSISKPTLNTSLEQLLRGVSVINNTANN-RIVVNS 510  
QY 522 IGSNKQIOVGA--DGIK-----FADVNVNVSNAAKFTTTRITEE 560  
Db 511 INLSNGLTLHTRKDGKVKINGDITSNENGNLTIKAGSWVDVHKNTITLGTGF-LNIVAGDS 569  
QY 561 IGFADADGKVKKSPYLDKQLQVGVKITKD-----SGINAGDOKISNVKDAT 609  
Db 570 VAF-EREG--DKARNATDAQITAQGTITVKNDDKQFRFNVSNLNGTGGKLFKFIANNF 626  
QY 610 DDTDA-----VYTKOLKOVQDQADGALQSFSIRDEKGEFTISNLSYNG----- 653  
Db 627 HRFEDGEINISGIVTINQTKDKVYWNASKDSYWNVSLLTNTVQKFTPIKFVDSGSCNQ 686  
QY 654 ---NPTNTFTTTFAGENGISNDIAKGVKVDIPINGLTTPKLTIVGSDKDGKTLV 710  
Db 687 DLSSRRSPAGVHFNGIGGKTFNFGANAKALFKLP-NAATDPK-----KELPITFNA 739  
QY 711 EQVASGNDTKNII-----RGLSPTLPSITNAGVVRTTEQGNITTSDEDKSKAASTG 761  
Db 740 NITATGNSDSSVDFDIHANLTSRAGINMDSINITGGL-----DFSITSHRNSNAFEIK 794  
QY 762 DIL-----NTGNLKN-----NSNSVGFVSTYNTVDFIDGNATTAKVTYDETNOTSKYT 810  
Db 795 KDLTINATGNSFSLKQTKDSFYNEYSKHAINSNHLTILGN-----VTLGGENSSSIT 849  
QY 811 YDWNDEK-TIETGNGKTNKIGVKTTLTNTNANGKATNFSTTDNDLVNAKDAEAL 869  
Db 850 GNINITNKANVTLOADTSNS-TGLKKRTLTGNSISVEGNLSLTGANANIVGNLSAED- 907  
QY 870 NTLAKEHTTKGPADTALQTFKVKGDGATDDTITVKGDTGONGKTVNPLKL---KGENG 926  
Db 908 -----STFKGASDNLAI-----TGTFNNGTANINIKGVVKLGDINNKGG 948  
QY 927 LVATNKGDTVTFINGTQSLKAGDSTTLNKDGLSTKNPASNEQIQVGADGVKFAKVDKG 986  
Db 949 LNIITNASGTOKIIN-----GNITKEKGLNKNKAKADAIIQIG-----G 989  
QY 987 NSSTGIDGTSRITKDOIIGFT-----GANGSLDITTKPHL---TKDKLVGVEVLT 1032  
Db 990 NISQK-EGNLTISDKVNITNQTIKAGVEGGSDESSEANLMTTQTKELLAGDLNLT- 1047  
QY 1033 NTGINAG---GKKITNTQSGDITONSND--VITGGRVYDLKTBLES---KINSAAKTAQN 1084

Db 1048 -SGFNKAEITAKNGSDLTIGNASGCGNADAKKVTDFKVKDSKISTDGHNVTLNSEVKTSG 1106  
QY 1085 SLHEFSVADEQGN-----HFTVSNPYSDVTSKTSKSDVITFAGENGITTKVNGKV 1133  
Db 1107 S-----SNAGDNSGCLTISAKDVTVNNVTSHTKNIS-----AAGNVVTTK--EGT 1152  
QY 1134 VRVGIDOTKGLTTPKLIVGNNGKGIIVDSKQGN-TITGLSNTLANVNDGAGHAJSQG 1192  
Db 1153 T---INATFG--SVEVTAONGTIKGNI-----TSQNVTVTATENL---VTTENAVINATSG 1200  
QY 1193 LANDTDKTRAASIGDVLNAGFNLOQNGEAVDFSVYDVIDGNATTAKTATVYDDTSKTS 1252  
Db 1201 TVNISTRT---GDIKGGIESTSGNVN-----ITASGNTLKVS----- 1234  
QY 1253 KVVYDVVNDKNTIETVSDKILGVKTT---TLTKTSANGNATKFSAADGDALVLRASDIAT 1308  
Db 1235 ---NITGQDVIVTAD--AGALTITAGSTISATTGNANIT-----TKTGDLNG 1276  
QY 1309 HLNTLAGDIO--TAKASOASSASVYDADGNKVIYDSTDKKYYQVNDKGQVKNKEVAKD 1367  
Db 1277 KVSESSGSVTLVATGATLAVGNIS-----GNTVTITTA-----DSGKLTSS----- 1315  
QY 1368 KLVAAQOTPDGTLAQMVNKVSVINKQVNDANKKQGINEDNAFIKGLNAAKDTKTNAAV 1427  
Db 1316 ---TVGSTINGT---NSVTSSQSGD-----IEGTISG---NTVNVTA 1349  
QY 1428 TVGDLNVAQAOTPLTPAGDTGTTAKKLGETLTIKGGQDTNKLTDNNIGVAVGTGFTVKL 1487  
Db 1350 STGD-----LTIGNSAKVEAKNGAATLTAESG-----KLTTQTGSSITSSNGQTTLT 1396  
QY 1488 AKDLTNLSVNAVAGVRIDEKGISFVDANGQAKANTPVLISANGLDLGGKRISNIGAAVDN 1547  
Db 1397 AKDSSIAIGNAANAANVTLLNTGT--LTTTGDSSKINA---TSGLTINAKDAKLDGGAAGDR 1451  
QY 1548 DAVNFKQFNEVAKTVNNLNNSNCSASLPFVVTDANGKPIINGTDCGPKOKAIG--ADGKY 1605  
Db 1452 TVVNATN---ASGSGNVTAKTSSSVNITGDLNTINGLNIISENGRNVRLRGKIDVKY 1507  
QY 1606 YHANANGVVPKDGKPIITDADKLANLAAGHPKLDAGHQVWASLG-----GNSDAIT 1656  
Db 1508 IQPGVASVEVETAKRV--LEKVKDLS-----DEERETLAKLGVSAVRFPVPPNNAITVN 1559  
QY 1657 LTNIKSLTP 1665  
Db 1560 TONEFTTKP 1568

## RESULT 12

US-08-617-697-9  
; Sequence 9, Application US/08617697  
; Patent No. 5977336  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/617,697  
; FILING DATE: 01-APR-1996



QY 1583 NCKPINTGKQKAIKG--ADGKYHANANCVVDKGPITDADKLANLAHCKPLDA 1640  
DB 1482 NGLIISGRNTRVLKREIDVKYIQGVASVEEVIEAKRV--LEKVKDLS-----DE 1533  
QY 1641 GHQVNASLG-----GNSDAITLTIKSTLP 1665  
DB 1534 ERETLAKGVSAVRFEPPNNAITVNTQNEFTTKP 1567

RESULT 13  
US-08-728-470-10  
Sequence 10, Application US/08728470  
Patent No. 5928651  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,470  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1529 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-728-470-10

Query Match 4.4%; Score 518; DB 2; Length 1529;  
Best Local Similarity 21.9%; Pred. No. 9.2e-21;  
Matches 389; Conservative 232; Mismatches 668; Indels 488; Gaps 87;

QY 97 LAKAHANQAIAGGKPPRNOAANQKAGSHAKGESIAIGGDVLAEGDASIAIGSDLY 156  
DB 2 MSVHTGATMQVDGNKTIIRN-----SYNAIINWKQFN 34  
QY 157 LDRN-----SYNSKYPNGLLSTLTONHTVLRQIRDSNGSQKYRTAAEGHASTA 208  
DB 35 IDQEMEQFLOESSNAVFNRVTSQDISQ-----LKGILDSNG-QVFLINP-----NG 81

QY 209 MAYAKGHFANAFG--TRSTAE--GNYSL--AVGLTAKAEKGYTIIAGSNAQAIIYNALALGAD 265  
DB 82 ITICKDALIINTNGTASTLDISNENIKARNFTLEQTOKDAL-----AEIVNHLITVGD 136  
QY 266 TRVDLYG-----IALGYSQILNNNNNNKAY---VPEGNSNIKSSKA 308  
DB 137 GSVNLIGGKVKNEGIVSNGGSIISLAGOKITTSIDIINPTTYSIAAEPENAINLGDIFA 196  
QY 309 TGNGLFSIGSSTIKRK-----IINVAGYEDTDVNVAAQLKAVENLAKR 352  
DB 197 KG-GNINVRAATIRNKGKLSADSVSKDKSGNIVLSAKEGEABEGVISAQOQAKG--GKL 254  
QY 353 QITFKGDDNGTGVKKKLGETLTI---KGETO---ADKLTDNNNIGVYTDNNNTGLKVKLAK 407  
DB 255 MIT-----GDKVTLKTCGAVIDLSGEGGETYLGDERGEGCN-----GIQLAK 297  
QY 408 NLSGLEIVSTKNLTASEK-----VTVSGSNNTAE---LOSGLLTFTPTTNASTDK 454  
DB 298 KTT-LEKGSTINVSGKEKGGRAIYVWGDIALDGNINAGGSDIAKTGGFVETSGHDSIGD 356  
QY 455 TVYGTGDLKFTDNSNTALEDTT---RITKDKIGFSNKAQTVDE-----NKPYLKDKL 505  
DB 357 DIVYDAKEMLLDDPDVSIETLTSGRNNITGENQVTTGDCTKESPKGNSISKPTLTNSTLE 416  
QY 506 -----VGNSTLANGGLTVNNTGGSNKQIOVGA--DGIK----- 537  
DB 417 QILRRGSYVNITANN--RIYVNSSINLSNGSLTHTKRDGVKINGDITSNENGNITIKAGS 475  
QY 538 PADVNVVNSNAAKFGTTRITEEIGFADADGKVDKPSYLDKQLQVGVKITKD----- 592  
DB 476 WDVHKNITLTGTG--LNIVAGDSVAF--EREG--DKARNATDAQITAQGTIIVNKDDKQFR 531  
QY 593 -----SGINAGDQKISNVKDATTDTDA-----VTYKOLKOVQODADGALQSP 634  
DB 532 ENNVISINGTGKGLKFIANQNNFTHKFCDELATISGIVTINQTTKKDKVYNWASKDSYWNVS 591  
QY 635 SIRDEKGOEFFISNLYSNG-----NTPNTPETITFAGENGISISNDIAKGVKVGIDP 687  
DB 592 SLTLNTVQKFTFIKFDVSGSNSQDLRSSRRSPFAGVHFNIGGKTFNFCANAKALFKLKP 651  
QY 688 INGLTTPKLTGVSKDKGTQLVIEQVASGNDITKNI-----RGLSPTLPSTTNAGG 738  
DB 652 -NAATDPK-----KELPITFNANITATGNSDSSVDFDIHANLTSAAGINMDSNITGG 704  
QY 739 VRTTEGGNTITSDEDKSAAASIGDIL-----NTGFNLKN-----NSNSVGFVSNTYND 787  
DB 705 L-----DFSITSHRNSNAFEIKKDLTINATGNSFSLKOTKDSFYNEYSKHAINSSHNT 759  
QY 788 FIDGNATTAKVYDETNOTSKVTVYDYNVDEK--TIELTGDNGKTNKIGVKTTLTTTNANG 846  
DB 760 ILGGN-----VTLGGENSSSITGNINITKANVTILOADTSNSN--TGLKRTLTILGNISV 813  
QY 847 KATNFSTTDNDALVNAKOTAENLNTLAKEIHTTKGADTALQTFVKKDKGATDDEITIVG 906  
DB 814 EGNLSLTGANANIVGNLSAED-----STFKGEASDNLNI-----TGFTT 853  
QY 907 KDGTONKTVN-----TLKLGKE-----NGLTVATNKDGTVTFGINTQSGLKAGDSTTLN 958  
DB 854 NNGTAN---INIKQGVVKLQGDINNKGGNLNITNAGTOKTILN-----GNITNEKG 902  
QY 959 GLSIKNPASNEQIQVADGVKFAKVDKGNSSSTGIDGTSRITKQDQIGFT-----G 1007  
DB 903 DLNINKIKADAETIQG-----GNISQK--EGNLTISSDKVNITNQITIKAGVEGG 950  
QY 1008 ANGSLDTTPHL---TKDKLVGEVEITWGINAG---GKKTINIQSGDITONSNDAA--V 1059  
DB 951 RSDSSAEANLTIQTKELKLAGDLNI--SGFNKAIEITAKNGSDULTIGNASGNADAKV 1008  
QY 1060 TGGRVVDLKTLEBS---KINSAAKTAQNSLHEFSVADEQCN-----HFTVSNPY 1105  
DB 1009 TFDKVKDSKISTDGHVNLNSEVKTSGNS-----SNAGNDNSTGLTISAKDVTVNNNV 1061  
QY 1106 SSYDTSKTSVITTFAGENGITTKVKNKGVVRVIGIDQTKGLTTPKLTVGNNGKGIIVDSKD 1165

Db 1062 TSHKTNIS-----AAAGNVTRK--EGTT---INATG--SVEVTAQNGTIKGN-----T 1105  
QY 1166 GON-TITGLSNTLANVTDGAGHALSOGLANDTKTRAASIGSDVNLNAGFNLCQNGEAVDF 1224  
Db 1106 SQNVTVATENL---VTENAVINATSGVNIKT-----GDIKGIESTSNVN----- 1153  
QY 1225 VSYDVDFJDNATTAKVYDDTSKTSKVYVDVNDKTIETVSDKKLGKVT-----TL 1280  
Db 1154 -----ITASGNTLKVS-----NITGQDVTVTAD--AGALTTAGSTI 1188  
QY 1281 TKTSANGNATKFAADGDALVKASDIATHLNTLAGDIO-TAKASOASSASVVDAGNK 1339  
Db 1189 SATTGNANIT-----TKTCDINGKVESSGSVTLVATGATLAVGNIS-----GNT 1233  
QY 1340 VIYDSTDYKYQVNDKQGVNDKNEKVAQKVAQAQTPDGTFLAQMNVKSVINKQVNDANK 1399  
Db 1234 VTITA-----DSGKLT-----TVGSGTINGT-----NSVTTSSQSGD--- 1265  
QY 1400 KOGINEDNAFIKLENAADTKTKNAAVTVGDLNNAVAQTPLTTFAGDTGTAKKLGETLTI 1459  
Db 1266 -----IEGTISG---NTVNVVTAAGD-----LTIGNSAKVEAKNGAATLTA 1303  
QY 1460 KCGQTDNKLTDNNIGVWAGTDFVTKLADLTNLNSVNAAGTTRIDEKGISFVDANGQAK 1519  
Db 1304 ESG-----KUTTGGSITSSNGQTTLTAKDSSIAIGNAANVTLNNTGT--LTTTGDSK 1356  
QY 1520 ANTPVLSANGLDGGKRISNIGAAVDNDNAVNFQKNEVAKTVNNLNQNSGASLPFVV 1579  
Db 1357 INA---TSGTLTINAKDAKLDGAASGDRVVNATN-----ASGSGNVTAKTSSSVNTIGDL 1409  
QY 1580 TDANGKPIGTDGKPOKAIKG--ADCKYYHANANGVPVDRKGPITDADKLANLAHAKRP 1637  
Db 1410 NTINGLNIISENGRTVRLGRKEIDVKYIQPGVASVEEVIEAKRV--LEKVKDLS----- 1462  
QY 1638 LDAGHQVWASLG-----GNSDAITLTNIKSTLP 1665  
Db 1463 -DEERETLAKLGVSARVFRPNNAITVNTQNETTKP 1498

RESULT 14  
US-08-719-641-10  
; Sequence 10, Application US/08719641  
; Patent No. 6218141  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/719,641  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-625  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1529 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-719-641-10  
  
Query Match 4.4%; Score 518; DB 4; Length 1529;  
Best Local Similarity 21.9%; Pred. No. 9.2e-21;  
Matches 389; Conservative 232; Mismatches 668; Indels 488; Gaps 87;  
  
QY 97 LAKAHANQATAIGSKRPDPNRQANQKAGSHAKGESIAIGGDVLAEGDSIAIGSDLY 156  
Db 2 MSVVHGTATMQVDGNKTTIRN-----SVNAINRWQFN 34  
  
QY 157 LDNR-----STNSKYPNGLLSTLQNIHTVLRQIRDSNGSKYRRTAAEGHASTAVGA 208  
Db 35 IDQNEHQFQLESNSAVFNRVTSQISQ---LKGILDSNG-QVFLINP-----NG 81  
  
QY 209 MAYAKGHFANAFG--TRSTAE--GNYSL--AVGLTAKAEKGYTTAIGSNAQAQINYGALALGAD 265  
Db 82 ITIGKDAIINTNGFTASTLDSINENIKARNFTLEQTKDKAL---AEIVNHGLITVTKD 136  
  
QY 266 TRVDLDYG-----IALGYCSQILANNNNNNKAY---VPEGNGSNIKSKA 308  
Db 137 GSVNLIGGKVKNEGVSIVNGSGSISLLAGQKITTSIDIINPTITYSIAAEPENAILGDI 196  
  
QY 309 TGNGLFSGSSTIKRK-----IINVAGYEDTDVNVVAQLKAVENLAKR 352  
Db 197 KG-GNINVRATIRNKGKLSADSVSKDSKGNIVLSAEGEAEIGGVISAQNOQAKG-GKL 254  
  
QY 353 QITFKGDDNCTGVKKLGETLTI---KGGETO---ADKLTNNNIGVVYTDNNTGLKVKLAK 407  
Db 255 MIT-----GDKVTLTKGAVIDLSGKEGGETYLGDERGEGKN-----GIQLAK 297  
  
QY 408 NLSGLEIVSTKNLTASEK-----VTVSGNNTAE---LQSGGLTFTPTTNASTDK 454  
Db 298 KTT-LEKGSTINVSQKEKGRAIVWGDIALIDGNAQGSIDIAKTGTFVETSGHDLISGD 356  
  
QY 455 TVYGTDLGKFTDNTALEDTT---RITKDKIGFSNKGATVDE-----NKPYLDKDKLK 505  
Db 357 DVIVDAKEMWLLDPDVSIEITLTSRNNNTGENQGYTGDGKESPKGNSISKPLTNTSLTLE 416  
  
QY 506 -----VGNSTLNNGLTVNNTIGGSNKQIQVGA---DGIK----- 537  
Db 417 QILRRGSYVNITANN-RIYVNSSINLSNGSILTLHTKRDGVKINGDITTSNENGLTIKAGS 475  
  
QY 538 FADVNVNVAKEGTTRITEEIGFADADGKVDKSPYLDKQKQVGGVKITKD----- 592  
Db 476 WVDVHKNIITLGTGF-LNIVAGDSVAF--EREG--DKARNATDAQITTAQGTITVNDKQFR 531  
  
QY 593 -----SGINAGDQKISNVKDATDDTDA-----VTYKOLQOVQODADAGLQSF 634  
Db 532 FNNVSVINGTCKGLKFFIANQNNFTHKFDGELNISGIIVINQTTKDKVYWNASKDSYNNVS 591  
  
QY 635 SIRDEKQGEFTISNLSNG-----NTPNTFTTITFAGENSISINDIAKGVKVGIDP 687  
Db 592 SLTLNTVOKFTFFIKFVDSGNSQDLRSSRSFAGVHFGGKTNFNIGANAKALFKLP 651  
  
QY 688 INGLTTPKLVGSKDKGTQLVIEQVAGSNDTKNII-----RGLSTPLTSITWAGG 738  
Db 652 -NAATDPK-----KELPITFNANITATGNSDSSVMFDIHANLTSRAAGINMDSINITGG 704



Patent No. 5549897  
GENERAL INFORMATION:  
APPLICANT: BARENKAMP, STEPHEN J  
APPLICANT: ST. GEME III, JOSEPH W  
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/038,682  
FILING DATE: 16-MAR-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: BERKSTRESSER, JERRY W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-293  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-038-682-2

Query Match 4.1%, Score 482; DB 1; Length 1536;  
Best Local Similarity 21.0%; Pred. No. 9.2e-19;  
Matches 394; Conservative 257; Mismatches 649; Indels 572; Gaps 94;

QY 1 MNHYKVFIFNKATGTFMAVACAKSHSGSSSTAGQVCGSSPVIRLTRVATLAILVICAT 60  
Db 1 MNKIYRLAFKSKRLNALVAVSELAR---GCDHST--EKGEKPARM-KYRHLALPLSAM 53  
QY 61 LINGSAYQNNSKIAFGTTGNNDNASASNEASTAIGSLAKAHANQAIAGGSKPDPNQA 120  
Db 54 L-----LSLGVTIPOSVLAS-----GLOCMQDVVHGTATMQVDGNKTIIRNSV- 96  
QY 121 NOKAGSHAKGKESIAIGGDVLAEGDASIAIGSDLLYLDRN-----STNSKYENGLLS 172  
Db 97 -----DA--IINWKQFNIDQNMVQFLOENNSAVFNRTS 130  
QY 173 TLIQNHVTLRQIRDSNGSKYRRTAEGHASTAVGAMAYAKCHAFANAG-TRSTAE-GNY 230  
Db 131 NQISQ---LKGILDSNG-QVFLINP-----NGITIGKDAIINTNGFTASTLDSIS 177  
QY 231 SL-AVGLTAKAEKGVYTIAGSNAQAIIYAGALGADTRVDLDYG-----IA 275  
Db 178 NIKARNFTEQTKDAL-----AEIVNHCLITVGKGSVNLICGKVKNEGVI SVNGGSIS 232  
QY 276 LQYGOILNNNNNNKAY---VPEGNGSNIKSKATGNGLFSIGSSSTIKRK----- 324  
Db 233 LLAGOKITISDIINPTITYSIAAPENAVNLGDIKFAK-GNINVRRAATIRNOGKLSADSV 291  
QY 325 -----IINVAGYEDTDVAVNAQKAVENLAKRIFKGDNDGCTGVKKKLGFTLT- 375  
Db 292 SKDKSGNIVLSAKEGEAEIGGVISAQNOQAKG-GKLMIT-----GDKVTLTKTGAVIDLS 344

739 VRTTEGQNTIISDEKSKAAASIGDIL-----NTGFNLKN-----NSNSGVFVSTYNTVD 787  
Db 705 L-----DFSITSHNRNSAFRIKDLDTNATGNSFLSKQTKDSFYNEYSKHAIISSHNLT 759  
QY 788 FIDGNATTAKVYDETNQTSKVTVYDNNYDEK-TIELTGDNGKTKNGIKGVKTTTLTTLTTRANG 846  
Db 760 ILGN-----VTLGGENSSSITGINITNKANVTLOADTSNSN-TGLKRTLTILGNISV 813  
QY 847 KATNFSTTDNALVNKADIAENLTLAKEIHTTKGTADTALQTFKVKKDGATDDTETIVG 906  
Db 814 EGNLSLTGANANIVNLSTAE-----STFGEASDNLNI-----TGFTT 853  
QY 907 KDTQNGTKVN---TLKLKE---NGLVTATNKDGVTVTFGINTQSLKAGDSTLLNKD 958  
Db 854 NNGTAN---INTKQGVKLGQDNNKNGGLNIITNAGSTQKTIIN-----GNITNEKG 902  
QY 959 GLSIRNPASNEQIQVAGDGVKFAKVDKNSSTGIDGTSRITKDKQIGT-----G 1007  
Db 903 DLNINIKRADAIEIG-----GNISQK-EGNLTISSDKVNITNQITIKAGVEGG 950  
QY 1008 ANGSLDTTKPHI---TKOKLKVGEVEITNTGINAG---GKKTINIQSDITQNSNDA--V 1059  
Db 951 RSDSSEAEANATIQTKELKLAGDLNI--SGFNKAETAKNGSDLTIGNASGGNADAKV 1008  
QY 1060 TGRVVDLKTLES---KINSAAKQASLHEFSVADEGN-----HFTVSNPY 1105  
Db 1009 TFDKVKDSKISTDGHNVTLNSEVKISNGS-----SNAGDNSNCLTISAKDVTNNV 1061  
QY 1106 SSYDTSKTSDDVITFAGENGITTKVKNKGVYVIGDQTKGLTTPKLVGNNNGKGVIVDSKD 1165  
Db 1062 TSHKTNIS-----AAAGNVTK--EGTT--INATG--SVEVTAQNGTIKGNI---T 1105  
QY 1166 QON-TITGLSNFLANVTNDGAGHALSQGLANDTKTRAASIGDVLNAGFNLOGNCEAVDF 1224  
Db 1106 SONVTVTATENL---VTTENAVINATSGTVNISTKT-----GDIKGGIESTSGNVN- 1153  
QY 1225 VSYDTPVDIDGNATTAKVYDDTSTKSVYDVVNNKTIETVSDKILGVKIT---TL 1280  
Db 1154 -----ITASGNLTKVS-----NITGQDVTVTAD--AGALTITTAGSTI 1188  
QY 1281 TKTSANGNATKFSADGDALVKASDIATHLNTLAGDIO-TAKGASQASSASYVDADGNK 1339  
Db 1189 SATTGNNANIT-----TKTGDINGKVESSSGSVTLVATGATLAVGNIS-----GNT 1233  
QY 1340 VIYDSTDKKYQVNDKQGVNDKNEKAKKLVAQAQTPDGTTLAQMNVKSVINKEQVNDANK 1399  
Db 1234 VTITA-----DSGKLS-----TVGSTINGT-----NSVTTSSQSGD--- 1265  
QY 1400 KOGINEDNAFIKLENAKDKTKKNAVTVGDLNVAQTPPLTFAGDTGTAKKLGETLTI 1459  
Db 1266 -----IEGTISG---NPNVNTASTGD-----LTIGNSAKVEAKNGAATLTA 1303  
QY 1460 KGGOTDNTLTDNNIGVWAGTDGTVKLAKLDTNLNSVNNAGGTTRIDEKGISFVDANGQAK 1519  
Db 1304 ESG-----LUTTGSSIISSNGQTTLTAKDSSIAIGNINANVTLTGCT--LTTIGDSK 1356  
QY 1520 ANTPVLSANGLDLGGKRISNIGAAVDNDVAVNPKQFNEVAKTVNNI--NNOSNGSASLPFVV 1579  
Db 1357 INA---TSGLTITINAKDAKLDAASGDRTVNNATN---ASGSGNVTAKTSSSVNITGDL 1409  
QY 1580 TDANGXPINTGCKPKQAKG--ADGKYHYHANANGVYVDKDGKPIIDADKLANLAHGRP 1637  
Db 1410 NTINGLNIISNGRNTVRUGKEIDVKYIOPGVASVEEVTEAKRV--LEKVKDLS----- 1462  
QY 1638 LDAGHGVVASLG-----GNSDAITLTNIKSTLP 1665  
Db 1463 -DEERETLAKGVSAVRFVEFNNAITVNTQNEFTTKP 1498

RESULT 15  
US-08-038-682-2  
; Sequence 2, Application US/08038682



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:41:04 ; Search time 76.31 Seconds  
(without alignments)  
2309.893 Million cell updates/sec

Title: US-09-361-619-11  
Perfect score: 11694  
Sequence: 1 MNHLYKVFKNATGTFMVA.....NGSADTOGHVGAAGVGFHF 2314  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1222.5	10.5	2059	D82671	surface protein XF
2	953	8.1	1588	A86036	probable adhesin z
3	743.5	6.4	1190	A82615	surface protein XF
4	668.5	5.7	5188	B85547	probable RXF famil
5	642.5	5.5	2249	A41477	190K surface antiq
6	571.5	4.9	2340	B71704	cell surface antiq
7	560	4.8	3029	S76109	hypothetical prote
8	552.5	4.8	2020	C48399	ABC-type transport
9	546	4.7	4919	T31105	hypothetical prote
10	546	4.7	2514	F81045	hemagglutinin/hemo
11	536.5	4.6	5627	C83339	hypothetical prote
12	520.5	4.5	13288	T03099	mucin, submaxillar
13	511	4.4	2468	A83412	hypothetical prote
14	504.5	4.3	2660	E85822	probable invasiv z
15	504.5	4.3	2703	H81193	hemagglutinin/hemo
16	504	4.3	4152	T31102	filamentous hemagg
17	483	4.1	1536	A43855	high-molecular-we
18	472	4.0	1651	JC1340	outer membrane pro
19	471	4.0	1477	B13855	high-molecular-we
20	463.5	4.0	2383	D64962	probable membrane
21	463.5	4.0	4199	S76412	hypothetical prote
22	462.5	4.0	2055	T31110	extracellular matr
23	459.5	3.9	2893	A64556	toxin-like outer m
24	450.5	3.9	1577	A35140	hemolysin A precu
25	448.5	3.8	2232	T34434	hypothetical prote
26	448	3.8	3263	E82410	hypothetical prote
27	448	3.8	3591	S21010	filamentous hemagg
28	446.5	3.8	3300	D70575	probable PFE prote
29	445.5	3.8	1643	D71630	outer membrane pro

30	444	3.8	3194	2	D71917	toxin-like outer m
31	442	3.8	3535	2	E83641	probable hemagglut
32	440.5	3.8	1270	2	E85649	hypothetical prote
33	439.5	3.8	1965	2	S75200	fat protein - Syne
34	438.5	3.7	1645	2	JN0896	crystalline surfac
35	436.5	3.7	2399	2	H71879	toxin-like outer m
36	433.5	3.7	2902	2	C71953	toxin-like outer m
37	425.5	3.6	3016	2	S77300	hypothetical prote
38	425.5	3.6	3442	2	E82589	hemagglutinin-like
39	425.5	3.6	3455	2	B82519	hemagglutinin-like
40	425	3.6	2154	2	F83068	hypothetical prote
41	425	3.6	4688	2	F82885	hypothetical prote
42	424.5	3.6	4558	2	C82199	RFX toxin RxA VCL
43	424	3.6	1403	2	T17372	plasma membrane-as
44	419	3.6	2529	2	B64635	toxin-like outer m
45	418	3.6	1975	2	B81192	hemagglutinin/hemo

ALIGNMENTS

RESULT 1  
D82671  
surface protein XF1529 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: D82671  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: D82671  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2059 <SIM>  
A:Cross-references: GB:AE003982; GB:AE003849; MID:99106554; PIDN:AAF84338.1; GSPDB:GN  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrato, D.M.; Carrer  
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krueger, J.E.; Kuramae, E.E.; La  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF1529

Query Match	10.5%	Score	1222.5	DB	2	Length	2059
Best Local Similarity	24.0%	Pred. No.	3.2e-39				
Matches	599	Conservative	368	Mismatches	855	Indels	679
Gaps	115						
QY	39	GSPFVRLTRVATLAILVIGATLNGSAYAQNNKSIARFTGTGNNNDNASN-EASIAIGSL	97				
DB	13	GKEPLHTILEVATLSLSPYTEKANAOYVI--NDGMDGQCQRIVDNGSASGVETVA----	66				
QY	98	AKAHANQAIIGSK-PDPNRQANQKAGSHAKCKESTAIIGDVLAEADSIATIGSDOLY	156				
DB	67	TQCESEPTWTGTARFPGSGTAEQOGAS-----RNLTLGGSLYVN---SGQGVGVNDV-	117				
QY	157	LDNRNSTKYPNGLLSLIQNTLQIRTRDSNGSKYRRTAAEGHASTAVCA--MAYAKG	214				
DB	118	-----LNKTYTSIRMGSVITWN-TVA-----GTNATAIGSAQSSAADALKASLATKASG	164				
QY	215	HFANAFTRSTAEGNYSVLAVGLTAKBKG-YTIIAGSNAQAINYGALGADTRVDL-DY	272				

Db 165 ARAIAKAGADGVDYVALGSGTAGTGASSIALGINASAVN-CAVAVGGGALVTPDQ 223  
 QY 273 GIALGYSOILNNNNNNNNKAYPEPGNSNKKSKATGNGL--PSIG---SSTIK-RKII 326  
 Db 224 AVALGLNSVA---SFGKGLSGYDPKTKTTSTDSAAAMKSTLAAVSIGDVSNTLKRQLS 280  
 QY 327 NVGAGYEDTDVAVNAQLKAVENLAKR--QITFGKDDNCT-----GVKKKLGETLT 374  
 Db 281 GLAAGTSNTDAVNAQLKVDVEIASRGNLTSAGANSNVAPGSSVDLKNKTDLNLTITKA 340  
 QY 375 IKGGETQ-----ADKLTNNNTIGVVVD-----NNTGLKVKLAKNLSGLETVST 417  
 Db 341 TGSNDVQNLNKDKVKVTLAVGCDALLNTDGLALGTDVSLSTGLAI-----TDGPAVTA 394  
 QY 418 KNLTAASEKVTVGSGNNTAELSGGLFTPTTNTASTDKTVYGTGDLKFTDNTALEDTTR 477  
 Db 395 SGIDAGSKV-----ISHVAAGAVSET-----STDAVNGSOLNAVQVQASOP 435  
 QY 478 ITKDKLIGPSNKACTVDENKPYLDKDLKVGNSTLNNGLTVNNTIGGNSN-KOIQVGADGI 536  
 Db 436 VT-----FTGNEGAV-----KRLGQSVTVISGESSTAGTYSGGNLSKVDEAAG- 479  
 QY 537 KPADVNVVSNAAKFTTRITEEIEGFADADGKVDKSPYLDKKQLQVGGVKITKDSGIN 596  
 Db 480 ---RIHLQADSPKFGNVINN-----GG-----KISGVT 506  
 QY 597 AGDQKISNVKDATDDTDVATYKQLKQVQ-----QDADG 629  
 Db 507 AG-----TEETDAVNFSLKSISTAVDQGWTLTASGNSGSKVAGSGGTVDLKNKTDG 556  
 QY 630 ALQSPSIRDEKGOEFTISLNS-NGNTPNFTFTITFAGENGISINDIAKGVKVGIDPI 688  
 Db 557 NLTKSGSDSNDVFNLSKDFKVDGWTSGT-----TVNNDGKVGSDVALGTTGLTITD 612  
 QY 689 NGLTTPKLVGSDKDGKQTLVIEQVAGS---NDTKNITRGLSPTLPSTNACGVFTTQ 745  
 Db 613 PAVTASGIDAGSK-----VISHVAAGVYVSETSDAVNG-----SOLNAVQVQAS-QP 658  
 QY 746 NITSPEDSKAASIGDIINTGFLNKNNSVGFVSTYNTVDFIDGNATTAKVTYDETNO 805  
 Db 659 VFTFGNEGAVKRS-----LQOSVVISGESSTAGMYSGGNLSKVDEAAGRIHLQADSPK 713  
 QY 806 TSKVTYDVNVDEKTIELTDGNGKTKNI-----GVKTT-----TLTTNANGKATNFSTT 854  
 Db 714 FGNV--INNGKISGVTAGTEETDAVNFSQLKSISTAVDQGWTLTASGNSG-----SKV 766  
 QY 855 DNDALVNAKDIAENLNTLAKEIHTTKGTADTALQTPKVKKDGATDDEFTITVGKDGTONK 914  
 Db 767 ASGTTVDLKNKTDGNTL-----ISKSGSDSNDVFNLSKDFKVDGM 805  
 QY 915 TVNTLKLKGNGLTVAATNKDGTVTFEGINTQSGLKAGDSTTLNKGDLISKNPASNEQIQVG 974  
 Db 806 T-----SGTIVVNN-----DGKVGSDVALGTTGLTI-----832  
 QY 975 ADGVKPAKVDKGNSSGTGIDGTSRITKDDQIGFTGANGSLDTPKHLTKDKLKVGEVEITWT 1034  
 Db 833 -----TDGPA-----VTAS 841  
 QY 1035 GINAGGKLTINOGDITONSNDVAVTGGRYVDLTKTELESKINSAAKTAQNSLHEFSVADE 1094  
 Db 842 GIDAGSKVISHVAAGVYVSETSDAVNG-----SOLNAVQVQ-----878  
 QY 1095 QGNHFTVSNPYSDYTSKTSDDVTTFAGENGITTKVKNKVYRVGIDQTKGLTTPKLTJVGN 1154  
 Db 879 -----SOPVTFG-----NEGAVKRLGQSVVISGESSTAGTY 911  
 QY 1155 NGKCI--VIDSKGO--NTITGLNTLANVINDGAGHALSOGLANDTKTRAA-----SI 1205  
 Db 912 SGGNLKSVVDEAAGRIHLQADSPKFGNVINNGKII--SGVTAGTEETDAVNFSQLKSI 969  
 QY 1206 GDVLNAGFNLOGNEAVDFVSTYDTVDV--IDGNATTAKVTYDDTSKTSKVYVDVN-----1259

Db 970 STAVDOGWTLTASGANGSKVASGGTVDLKNKTDGNLTISK-----SGDSNDVFNLSKDFK 1024  
 QY 1260 -----VDNKTTEVTSDDKLGKVTKTTTLTKTSANGNATKFSAADGDALVKASDIATH 1309  
 Db 1025 VDMGTSGTTVVNNDGVKVGSDVALG--TTGL--TIANGPAVTASGIDA-----GSKVISH 1075  
 QY 1310 LNTLAGDIOTAKGASOASSASYVDADGNKVVIYDSTDKKYYQVNDKQGVNDKNEKVAKDKL 1369  
 Db 1076 V-----AGAVSETSDAVNGSOLNAVQVQASOPVTFGTGNE-GAV--KRLSGOSV 1123  
 QY 1370 VAAQTPDPGTLAQMNVKSVINKE-----QVNDANKKQGINEDN-AFIKGLENAAKDTKT 1422  
 Db 1124 ISGESSTAGTYSGGNLSKVDEAAGRIHLQADSPKFGNVINNGKISGVTAGTEETD- 1182  
 QY 1423 KNAAVTVGDNLNAVQATPLTEFAGDTGTTAKKLGETLTIKG-----GOTDTNKLTDNNI 1474  
 Db 1183 ---AVNFSQLKSI-----STAVDOGWTLTASGANGSKVASGGTVDLKNKTDGNL 1227  
 QY 1475 GVAAGTQDFTV--KLAKDLTNLNSVAGGTTRIDEKLSIF-----VDANGQAKANTPVLSA 1527  
 Db 1228 TISKSGSDSNDVFNLSKDF-KVDGMTSGTTVVNNDGVKVGSDVALGTTGLTIANGPAVTA 1286  
 QY 1528 NGLDLGCKRRTISNIGA-AVDD--NDAVNEKOFNEVAKTVNMLNNSGASLFPFVVTDA 1584  
 Db 1287 SGIDAGSKVISHVAAGAVSETSDAVNGSQLNAV-----QVQASOPVTF-- 1331  
 QY 1585 KPINGTDGKPKQKAI-----KGADKYYHANANGVPVDKDGK---PITDADKLANLA 1632  
 Db 1332 ---GNEGAVKRLGQSVVISGESSTAGTYSGGNLSKVDEAAGRIHLQADSPKFGNV 1387  
 QY 1633 AHGKPLDAGHOVVASLGG--NSDAITLTNLIKSTLPQIDTPTNGNANAGAOQLPSLSAAQ 1690  
 Db 1388 ---INNGKISGVTAGTEETDAVNFSQLKSISTAVD-----QGWTL-TASGAN 1431  
 QY 1691 QSNAA--VKDVLNVGFNLQTNHNVQVDFKAYDITVFNVNGTGADITSVRASDGTMSNITV 1748  
 Db 1432 GSKVASGGTVDLKNKTDGNLTISK-----SGDSNDVFNLSKDFKESITV 1476  
 QY 1749 -NTALAAATDDGNGVLIKAKDGKFKYKADLLMPNGSLKAGKSASDAKT-PTGLSLVNPNAK 1806  
 Db 1477 GNTQL---DKDG---VKYSSNVLLDSNEL---ITSHSSTSSKVLANGEVSVNRTVVN 1526  
 QY 1807 G---STGDAVALNLSKAVFKSKDGTGTTTSSDGIISITQGDNSISITLSKGLNVGKVI 1863  
 Db 1527 GDGVNIDVVVNDL-----GLSIVG--GASLTLS--GINAGSHKI 1563  
 QY 1864 SNVGRGTGTDTAANYQOLNEVRNLLGLG-NAGDNADGNQV---NIADIKK-DPN---SG 1915  
 Db 1564 TNVTAGTEDTDAVNFSQLKSIVSEAVDKGWTLTASGANGSKVYSGGTVDLKNKTDGNL 1623  
 QY 1916 SSSNKTVI-----KAGTVLGGKGNNDTEKLTATGGVGVVDKDGNGANG---DLSNVV 1964  
 Db 1624 SGDSNDVFNLSKDFKVDVETAG-----NTVVNTDGVKVGSDVSLGAMGLFIANGPSVTA 1678  
 QY 1965 KTKDGSKKALLATYNAAGQTNLTNNPA---EATDRINEQGIREFHVNDGNQEPVQGR 2021  
 Db 1679 SGFNAGDK---VISHVAGMADTDNAVNSQLKQAVQSVTVKATRYISTNDG---TOGG 1731  
 QY 2022 NGIDSASGKHVAIGFONKADGEAAVIGRQTA-GNOSTAIGDNAQATGOSTAIGTG 2080  
 Db 1732 NYDGDGATGSKAIAACVGTQASGEGAAVVGSAAGSGKSGTAIGRNAITASADGVALGDG 1791  
 QY 2081 ---NVWTGKHSG-----AIGDPSTVKADNSYSVGNNNQFIDAT-----2115  
 Db 1792 AKDGGGAESEYTGKISGVONNTVGVSDAAKGETRSISNVADAKEAMDVAVNLQDLAV 1851  
 QY 2116 -----Q-----DVEFGV--GNNITVTE---NSVALGNSAISAGTHACTQAK 2153  
 Db 1852 AOKSNLQTTDDMRHEINNIEDVFKITGDSASSVKGNGVNAAGITNAAVS-GTESVALG 1910  
 QY 2154 KSDGTAGTTTTAGATCTVKGFAGQTAAGVAVSVGASGAERRIQNVAAAGEVSATSDAVNGS 2213  
 Db 1911 NTNVSADNNAVAG-NGSVADRA-----NSVSGSGSERQVTVNVAAG---TADTDVNV 1961

Qy 2214 QLYKATGCIANATWELDHRIHQENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNG 2273  
||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 1962 QL--NQCLITAKQYTDGMVGNLRRETSGGVAAAIAIANLPQAYVQGRGMTSCGVSSVQG 2018

Qy 2274 QGAVAVGLSKLSDNCQWVFKNIGSADTQGHVGCAAVGAGGFHF 2314  
| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 2019 QSAAIVGSAVSSEGSCHWVKFPGSANTRSHVGVGAGVGYOW 2059

RESULT 2  
A86036  
probable adhesin Z5029 [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: A86036  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamouitis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: A86036  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1588 <STO>  
A:Cross-references: GB:AE005174; MID:g12518349; PIDN:AAG58749.1; GSFPDB:GN00145; UWGP:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z5029

```

Query Match      8.1%  Score 953;  DB 2;  Length 1588;
Best Local Similarity 21.8%;  Pred. No. 4.6e-29;
Matches 535;  Conservative 273;  Mismatches 645;  Indels 996;  Gaps 106;

Qy      1  MNHVIKVFKNKATGTFMAVAEACAKSHSGSSSSTAGOVGSPVIRLTRVATPLAILVIGAT 60
Db      1  MNKIFKVIWNPATGNTVYTSETAKSR--GKKS-----GRSKLL-----ISALVAGGM 45

Qy      61  LNSGAYAQNNKIAFGTTGN--NDN-----ASASNEASTAIGSLAKH-----ANQAI 106
Db      46  LS-----SFGALANAGNDNGOVGYDGSAGDGWVAIGKAKANTFMTNSGST 94

Qy      107  AIG-----GSKDPNRQAQNQRAGSHAKKESIAIGGDVLAEGDIAIATGSD 153
Db      95  AVGYDAIAEGQYSSAIGSK-----THAIGGSAWAFGVSAISEGDRSIALGAS 141

Qy      154  DLYLDNRST-----NSKYPNGLLSTLIQNHTVLRQIRD-----SNGS 190
Db      142  SYSIGQYSMALGRYSKALGKLISIAMGSSRAEGANAIALGNATKATEIMSIALGDTANAS 201

Qy      191  QKY-----RRTAAEGHASTAVGAMAYAKGFHAFGRSTAEAGYSLAVGLTAKAEKGYT 245
Db      202  KAYSMAWGASSVASEENAIAGAETEAEN--ATAIGNNAKAKGTTNSAMWFGSLADKQVNT 260

Qy      246  IAGSNQAQINYGALAIAGDPRVDLDYGAIALGYGSOILNNNNNNNNKAYYPEGNGSNIKS 305
Db      261  IALNGSOALADNAIAGQGNKADGVDYAIALNGNSQSRGLNTIALGTASNATGDKSLALG 320

Qy      306  SKATGNGL-----FSGISSTPIKKIINV--GAGYEDT--DAVNVQAOLKAV 346
Db      321  SNSSANGINVALGADSTADLDNTVSVGNSLSLKKRVNVKNGAIKSDSYDAINGSQLYAI 380

Qy      347  -ENLAKROIPTKGD-----DNGTGVKKKLGGETLTIKGETQADKLTDNNNIG-----VYTDN 397
Db      381  SDSVAKR-----LGGGAADVDDGDT---VTAPTYNLKNG-----SKNNVGAALAVLDE 425

Qy      398  NTGLKVKLAKNLSGLETVSTKNLNTASBKVTVGVSGNNTAEQLQSGGLTFTPTTNASTDKTVY 457
Db      426  NT-----LQWDQTKG-----KYSNAHG-----TSSPTASVITDV--- 454

Qy      458  GTDGLKFTDNSNTALEDTTRITKDKIGFSKNAGKATVDENKPYLDKDKLKVGNSNTLNNGLT 517
Db      458  -----LQWDQTKG-----KYSNAHG-----TSSPTASVITDV--- 454

```

Db	455	ADG-----TISASKDAVNGSQLKATNDD-----VEANTANIANTNSN-----	492
Qy	518	VNNTIGSNKQIQVGADGKIFADVNVVNSAAKFGTTRITE--BEIGFADAGDVKKKSP	575
Db	493	-----IATNTANIAT-NTTNTNLTSVGLQADALLWNET--	527
Qy	576	YLOKKQLQGVGKWTIKDSGINAGDQKISNVKDA-----TDDTDAVYVKQLKVQOQDADGALQ	632
Db	528	--KKAFAAAGQDT-----TSKITVWKDADLTADSTDAVNGSQLKTTNDAV--ATN	574
Qy	633	SFSTRDEKGEFTISNLYSN-GNTPNPTFTTIFACENGISISNDIAKGVKRVGIDPINGLE	691
Db	575	TTNIAN-----NTSNIANTNINISLTTETVNLGEDALK-----	608
Qy	692	TTPKLTVGSKDKGKTOLVIEQVAGSGNDTKNIIRGLSPLTSPITNAGGVRTTEQGNITISD	751
Db	609	-----WDKD-----NGVFTAAGH-----	621
Qy	752	EDKSKAASIGDIINTGPNLKNNSNSVGFVSTYNTVDFIDGNATTAKVYDYDTNOSTKVTY	811
Db	622	-----TETTSKIT-----	629
Qy	812	DVNVDEKIELTGDNKTNKIGVKTTTLTTTNANGKATNFSTTDDNDALVNAKDAENLNT	871
Db	630	--NVKD-----LTTGSTDVAVG--SOLKTTNDVATNTNIATNTN	669
Qy	872	LAKIEHTKTADTALQOTFKVKKDGATDDTITVCKDGTQNGKTVNTLKLKGENGLTVAT	931
Db	670	ISNLTEVTNLGEDALKWKD-----DNGVFTAAGHNTASKITNL-----	710
Qy	932	NKDGTVFGINTOSGLKAGDSTTLNKGDLISIKNPASNEQIQVGADGVKFAKVDKGNSTG	991
Db	711	--DGTVT-----ATSSDAIN--GSOLYDLSLSNIATVFG-----GNASVN	745
Qy	992	IDGTSRITKDOIGFTGANGSLDTRKPLTKDKL-----KVGEVEITWT-----	1034
Db	746	TDGY-----FTGPTYKIGETNYNVGDDALAAINSFSTLSGALLMDATAGKFSAK	796
Qy	1035	-GINAGGKKTINTQSGDITQNSNDVATGGRVYDLKATELESKINSAAK--TAQNSLHEFSVA	1092
Db	797	HGTNGDASVITDVADEIGDSSSDAVNGSQLHGVSVYVDALGGGAENVADGTTATPYT	856
Qy	1093	DEQGNHFTVSNPYSSYDTSKTSVITF-----AGENGITTKVKNKGVVYRGIDTKGLTTPKL	1149
Db	857	IANADYDNVGDALNAIDTT-LDDALLMDADAGENAFSAAH-----GKDKTASVIT--	906
Qy	1150	TVGNNGKGVVIDSKDQNTITGLSNVLANTNDGAGIALSOGLANDDTKTRAASIGDVL	1209
Db	907	NVANGAISAASSDAINGSQLYT-----TNKYIADALGDGAENVADGTTATPYTIA	957
Qy	1210	NAGFNLOGNEAVDFYSTVDYDFIDGNATTAKVYDDTSKTSKVYVDVNDKNTIEVTS	1269
Db	958	NAEYNNVG-----DALDALDDNA-----LWDETAGNGAGAYNASHDGKASIIIN	1002
Qy	1270	DKKLGKVTMTTLTKTSANGNATKFS--AADCDAVLKASDIATHLNT-----LAGDIOQAKG	1322
Db	1003	-----VANGSISEDSTDVANGSQL-NATNMWIEQNTQIINQLAG-----	1040
Qy	1323	ASQASSASVADAGNKVYDSTDKKYQVNDKQGVDDKNKEVAKDKLVAQAQTPDGTIAQ	1392
Db	1041	-----NTDATYIOENAG-----INYYRTNDGDLA--	1065
Qy	1383	MNVKSVINKQVNDANKKOGINEDNAFICKLENAAKDTKNAAVTCVDLNAVAQTPLTF	1442
Db	1066	-----FNDAS--AGCV--GATAIGYNSVAKG--DSSVALIGQGSYSDV-----	1101
Qy	1443	AGDTGTTAKLIG----ETULIKGGQDPTNKLTDNNIGVVAG----TDGFTVKRLAKDLTNLN	1495
Db	1102	--DTGIA--LGSSSVSSRVIKAGSRDTS-ITEN-GVVYGYDITDG-----ELLGAL	1146
Qy	1496	SVNAGGTTRIDEKIGISFVDANGQAKANTPVLNGLDLGCKRISNIGIAAVDNDVANPKQF	1555
Db	1147	SIGDDGKY-----ROINVADGSEAHDAV-----	1170

Qy 1556 NEVAKTVNNLNQNSGASLPFFVVTDDANGKPIGTGDKPQKAIKAGDCKKYHHANANGVPV 1615  
Db 1171 ---TVPOLNATCAVATTP-----TKYFHANS----- 1194  
Qy 1616 DKDKPITDADKLA-----NLAHAKPLDAGHQVVA-SLGGNSDAITLTNTKSTLPQIDTP 1670  
Db 1195 -----TEEDSLAVGTDSLAKGAKTIYNGDKGIGIGYAGVVDANALNGIA----- 1238  
Qy 1671 NTGNANAGQAQSLSLAAQNSNAASVKDVLNVGNLQTNHNVDFVKAYDTVNFVNGTG 1730  
Db 1239 ---IGSNAQVIH-----VNSIAIGNS- 1257  
Qy 1731 ADITSVRSADGTMSTNITVNTALAAATDDGNVLKAKDKFKYKADDLMPNGSLKAGKSASD 1790  
Db 1258 ---TITRCAQ----- 1264  
Qy 1791 AKTPTGLSLVNPNGAGKSTGDAVALNLSKAVFKSKGTTTTVSSDGISTQGDNSISIT 1850  
Db 1265 ---TNTAYNNDAPQNSVGE-----FSVGSAD----- 1288  
Qy 1851 LSKDGLNVGGKVIQSVNGKGTWTDAAANVOQLNEVRNLLGLGNAGDNADGQVNIADIKK 1910  
Db 1289 -----GQRIITNVAAGSADTDVAVNGQL-----KVTDQV 1318  
Qy 1911 DPNSSSSNRRTVIRKAGTVLGGKGNNDTEKLGATGGVQVGVKDGNGNGLSNVWVKTKQDG 1970  
Db 1319 SONTOSITN-----LDNRVTNLDNRVT-NTENGI-----GDIV----- 1350  
Qy 1971 SKKALLATYNAAGQNTYLTNPAEALDRINEQGRFHFVNDGNEPVPVQGRNGIDSSASG 2030  
Db 1351 -----TGSTKYKPTN-----TDGVDASAQ 1371  
Qy 2031 KHSVAIGFOARADCEAAVAICROTAQAGNSTAIGDNAQATGDSIAIGTGNVTVGKHSGA 2090  
Db 1372 KDSVAI-----GSGSIAADN-----SVALGTGSVAT----- 1398  
Qy 2091 IGDPTVKADNSYSGVNNNQFIDATQDFVGGNNITVTESNVALGNSAISAGTHAGT 2150  
Db 1399 -----EENTISVGSST-----NORRIT-----NVAAGKNA-----TDAVN 1428  
Qy 2151 QAK-KSDCTAGTTTAGATGTVKGFAGTAVGAVSVGASGAERRTONVAAGEVSNATSDA 2209  
Db 1429 VAQLKSSBAGVRVDTKADGSID--YSNITLG-----GNGGTTTRISNVSAQ-----VNNNDV 1479  
Qy 2210 VNGSOLYKATOGIANATN-----ELDHRIHONENKANAGISSAMAMASMPQAIYFGRSMVT 2265  
Db 1480 VNYAQLKQSVQETKQYTDQRMVMDNKLKSKTESLGGIASAMAMTGLPQAYTPGASNAS 1539  
Qy 2266 GGIATHNCGGAVAVGLSKLSDNGQWFKINGSADTQGHVGAAGVGFHF 2314  
Db 1540 IGGGTNGESAVAGLVSMVSGRWRVYKLOGSTNSQGEYSALGAGIQW 1588  
  
RESULT 3  
A82615  
Surface protein xfl981 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: A82615  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: A82615  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1190 <SIM>  
A:Cross-references: GB:AE004017; GB:AE003849; NID:g9107083; PIDN:AAF84783.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincanl, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La  
J.D.; Junqueira, M.L.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins  
Chado, M.A.; Madeira, A.M.B.N.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.  
A:Authors: Martins, E.M.F.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,  
F.G.; Nunes, L.R.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa  
Rodrigues, V.; Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Vojvovski-Almeida, S.; Vettore, A.L.  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF1981

Query Match 6.48; Score 743.5; DB 2; Length 1190;  
Best Local Similarity 24.1%; Pred. No. 3.3e-21;  
Matches 352; Conservative 188; Mismatches 481; Indels 441; Gaps 67;  
  
Qy 1010 GSLDTTPHLTKDKLVKCE-----VEITNTGINAGCKKITNTQSGDITQNSDAVTTGGR-- 1063  
Db 13 GGIEGVDP-LFLPKYKICOSLOHAVMTSA---ASSKK-----GTQPRRSNNAMTAKRSA 62  
Qy 1064 -VYD-----LKTELESKINSAAKTA-----ONSLHEFSVADEQGNHFTVSNPY 1105  
Db 63 KLHRRQLHIVLLTFLAASTGYTGKVAQVYVNSDSTENCVEILGDSQTSQTSFIHSASNDK 122  
Qy 1106 SSYDTSKTSDDVITFAGEN---GITTKNKGVVRVYCIDQTKGLT-TPKL-TVGNNGK--- 1157  
Db 123 CKPFTQTEYSLFYDRLVGLGSLYVNEG--KLGLVDISGATSMRLGSLTATNGSAGI 180  
Qy 1158 -GIVIDSKDGQNTITGLSNTLANVTNDG--AGHALSQGLANDTOKTRAASIGDVLNAGFN 1214  
Db 181 DSIAGSGQSGKT-----DQNTSGATVAOGL-----RSIAGTGTARS--- 217  
Qy 1215 LQNGEAVDFVSTYDVFDFIDGNATTAKVYDDTSKTSKVYDVNDKNTTEVTSDDKLG 1274  
Db 218 -----QSQDASIGTCASTTGNFAI----- 237  
Qy 1275 VKTTTLTKTSANGNATKPSAADGALVKASDIATHLNTLAGDIOTAKGASQASASVYD 1334  
Db 238 -----AIGNGALTSIANGALGASSVTR-----GGVALGQGSALATASGI--- 279  
Qy 1335 ADGNKVIYDSTDKYKYQVNDKQVKNKEVAKDLVAQAQTPDGTFLAQMNVKSVINKQV 1394  
Db 280 ---TGYPDPVTKS-----TSTLSTSMRSTLGVASI 306  
Qy 1395 NDANKQGINEDNAFIKGLNAAKDTKTNAAVTVGDLNVAQOT-----PLTFAGDTGTT 1449  
Db 307 CNITVSS---TSQTRQLTGLAAGRSQTD-----AVNVAQLKLLAESVGGWNLTASGANSN 359  
Qy 1450 AKKLGETLTIKGGOTDNTKLTNNINVGAVAGTDG--FTVKLAKDLTNLNSVNAAGGTTRIDEK 1507  
Db 360 V-ALGESVDLKN-----SDGNLLITKTTDSNDVTENLATAL-KVDSLATGTNTAMTDD 409  
Qy 1508 GISF-----VDANGQAKANTPVLSANGLIDLGCKRISNIGAAVDDNDVNFQFNVKATV 1562  
Db 410 GVTVGKRVTLDSGLTVAEGSPVSISSGGINAAAGQKIMNVTGTADTDVNFQOLQAVSDTA 469  
Qy 1563 NNLNNSOSNGASLPVPTVDANGKPIGTGDKPQKAIKAGDKGYIHANANGVPVDDKGP 1622  
Db 470 SKGWNLLASGYN-----SSNVAPCASVD-----LKNYDGNLLITKAIGI----- 508  
Qy 1623 TDADKLANLAHAGRPDLAGHQVAVSLGNSDAITLTNTKSTLPQIDTPTNGA-NAGAAQ 1681  
Db 509 --NDVTENLAT---ALEADSLTGTNTAMTDTGVTGS-----NVTLGSTGLVITDGPVS 557  
Qy 1682 SLPSLSAAQ---SNAASVKDVLNVGNFNLQTNHNVQDFVKAYDTVNF-----VNGTGADI 1733  
Db 558 TSSGISAGNQNKITNVAAGTADT-----DAVNFQOLQAVSSTASKG 597  
Qy 1734 TSVRSADGTMSTNITVNTALAAATDDGDNVLI-----KAKDGKFKYKADDLMPNGSLKAGKSAS 1789



```
Db 598 WNLASGANSNVVPGESVDLKNDSGLMLLTKTDSNDVTFNLATALKVD-SLTTGNT-- 654
QY 1790 DAKPTGLSLVNPAGKSGTDAVALNLSKAVFKSKDGTITTTVSSDGSIOGKDNSSI 1849
Db 655 -AMTIDGVT-VGSNVTLGSGT-----LVITDGPVS----- 682
QY 1850 TLSKDLNVLGCKVSVNGKTKTDAANVOOLNEVR-----NLGLG-NAGN----- 1895
Db 683 --TSSGISAGNOKITNVAAGTADTDAVNFQLOAVSSASKGNLLASGANSNVAPGES 740
QY 1896 ---DNADGNVNIADIKKDNSSG-----SSNRTVIRKAGTVLGG 1931
Db 741 VDLKNTDGNV-----TSKESGNDVLFNLSSLLKDLKLTVDGTVMTNGTVGSGVTLSG 796
QY 1932 KG--NNDTEKATGGVOGVGDK-----DGNANGDLSNV-VWTKQDKGSKKALLATYNAAG 1983
Db 797 MGLVITDGPVSTSSGINAGSQKITNVAAGTADTDAVNLSQLNTAMAGS----- 844
QY 1984 QTNLYLNPAEATDRINEQIRFEHVNDGNQEPVQVRNGIDSSAGKHSVAIGFOAKAD 2043
Db 845 -----GAKSVHYISTYDGG-----TOGGNTNGDGTGTRSIAGVGVTLAS 884
QY 2044 GEAAVATIGROTA-GNOSIAIGDNAOATGQSTAICTG-----NVVTGKHSG----- 2089
Db 885 AEGATVGSAAASGKSTAGRNAVASADGSVALGDGAKDGARGAESYTGKYSGLQNTT 944
QY 2090 -----AIGDPSTVRKADNSYSYGNNNQFIDAT---QTDVFG-----VGNNTI-----TVT 2129
Db 945 VGTIVSGDASKGETRIVSNVADAKEATDAVNLQLDRVAQDANRYVDNKNKTESLSEGTFFV 1004
QY 2130 ESNVALGNSA--ISAGTHA-----GTOAKKSDGTA-GTTTATAGATGVKGFAGOTAVG 2181
Db 1005 KVNSL---NNSATPIAAGVDATAIGVATASGADSIAMGNKASASADNAV-----AIG 1054
QY 2182 -----AVSVCASCAERRIONVAAGEVSATSDAVNGSOLYKATOGIANATNELDHR 2232
Db 1055 NHSVADRANTVSGSAGSERQVNTNVAAG---TADTDVAVNSQL---NQGLITAKQYTDGV 1108
QY 2233 IHONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGGAVAVGLSKLSDNGOWVF 2292
Db 1109 VGSILRRDTDGVAATAATNLPOAYIPGRGMTSVGVSSYRGQSAIAGVSSVSESGRWF 1168
QY 2293 KINGSADTQGHVGAAGVAGPHE 2314
Db 1169 KFSGSANTRSQVGGIGAGVGQW 1190

RESULT 4
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B85547
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85547
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GB:AE005174; MID:g12513368; PIDN:AA654838.1; GSPDB:GN00145; UWGP:206
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0615
```

Query Match 5.7%; Score 668.5; DB 2; Length 5188;  
Best Local Similarity 21.6%; Pred. No. 1.7e-17;  
Matches 606; Conservative 290; Mismatches 1017; Indels 895; Gaps 141;

```
QY 29 GSSSSTAGQVSSPVIRLTRVATLAILVIG-ATLNGSAYAQNNSKIAPGTTGNNDNASAS 87
Db 283 GGTYYTATVAGDGSWTTTTPAADLSVLRDGDTAQVASV-----STINGTASAT 331
QY 88 NEASI-----AIGSLA-----KAHANQAIAGGSKPDPNQAANKAGSHAKGRESI 134
Db 332 HAYSVDATATLAINTIATDDILNAAEAGNPLTISGSSTAEAGQTVTVTL-----NGV 384
QY 135 AIGGDVLAEGDASIAIGSDLLYLDNRNSTSKYNGLLSTLIQNHTVLRQIRDSNGSKYR 194
Db 385 TYSGSVQADGSWSVSLPTADL---SNLTASQY-----TVSASVSDKAG----- 424
QY 195 RTAAEGHASTAVGANMAYAKHEFANAFGTRSTAEGNYSLAYGLTAKAEKGYTIAIGSNAQA 254
Db 425 -----NPASANHGLAVDLTVPV---LTINTVSGDDI 452
QY 255 INYGALALGADTRVDLDYGIALGYGSOILNNNNN-----NNKAFV-----PEGN 299
Db 453 INAA-----PHGQAL-----VSGSYGGGAGDVITVTLNSKYTYTMLDASGN 495
QY 300 GS-NIKSSKATGNGLFSIGSSTIKRKIIYNGAGYED---TDAVNVAQ----- 342
Db 496 WSVGVPAAVDTALG---SGPQTITAAITDAAGNSDDASRTVTNLAAPTIGINTIATDDV 552
QY 343 LKAVERNIAKQITFKGDDNGTGVKKLGETLITKGGTQADKLDNNNNIGVVTDNNTGLK 402
Db 553 IKATEKGADLQIT-----GTS-NQPAQTITV-----TLNGQNYTATTDSNGXWS 596
QY 403 VKLAKNLGSETVSTKNLTASEKVTGSGN-NTAE---LQSGGLTFTPTTNASTDKTVYG 458
Db 597 ATPV--ASAVSALGEANTVTANVTDTAGNSNSASHNVLNSALPAVTTINAVATDDLLINA 654
QY 459 TDGLKFTDNSNT-----ALEDTTRITKDKIGFSKNAGTVDENKPYLDKDKLVGNST 510
Db 655 AE---SGNAQTSIGQVTTGAAGDQTVTVT---LGGNIYATVQSNLSW-SVDVPAADIQA 706
QY 511 LNNGLTVNNTI-----GGSNKQIOVGA-----DGKIFADV----- 541
Db 707 LGNGDLTVNASVTVNGVNTGSGSRDITIDANLPLGRVDTVAGDDVINSIEHQALVITGS 766
QY 542 -----NVNVSNAKFGTTRITEEIECFADADCKVKKSPYLDKLOLVGGVITK 591
Db 767 SSGLTAGTALTVEINNTV-YGATVTL-----ADGTSVGLGVPADVSNMPAGTVNITV 816
QY 592 DSGINAGDQKISNVKDATDDTD--AVTYKQKQVQDADGALQSFISIRDEKQGQFTISNL 649
Db 817 -SGTNSAGTTSTIHPVTVDLAGVAITNL-----SGDDVINAV-----EKGETLVVSGS 866
QY 650 YSNQNTPTTETITTFAGENGISINDIAKVKVIGIDPINGLTTPKLTIVGSDKDKGTQOLV 709
Db 867 TSGVEAGQTV-TVTFGGKNYTTTVE--ANGSWTVNVPADLAALP-----DGAGNVQAS 917
QY 710 IEQVASGNDTKNIIRGLSPILPTITNAGVTRTTEGGNTITSDE-----DKSK 756
Db 918 VSNINGNSAADRAYSVDATAPLVTI-----NTIASDDILNVSEACAGITISGTT 967
QY 757 AASIGDILNTGFNLKKNNSNSGVFYSTYNTVDFIDGN-----ATTAKV 798
Db 968 TAQAQGLTV--TLNNN-----TYQTTVLADGTVSNVPAADLSGLTASSYTVTATV 1017
QY 799 TYDETNOTS---KVTVYDVNVDEKTIETLTGDKGNTKI-----GVKTTTITTT 842
Db 1018 SDRAGNPASADHALVVDITAPDLTINTVAGDDIINAJEHOALVVSSTGTSGAAAGDVTV 1077
QY 843 NANGKATNFSTNDALVN-----AKDI---AENLNTLAKEIHTTKGTADTALQTFVK 893
Db 1078 TLNCK--NYTTT-LDASGNWSVGIIPAADVATALATGQSTITASLDRAGNSDSTTHDVTVD 1134
QY 894 KDGAT-----DD-----ETITVKGDTQNGTKV----- 916
Db 1135 LSGPTLTINTVSGDDIINAAEIVVAQITISGOVGTAVAGNTVITVIGNOYNATVQSDLS 1194
QY 917 -----NTLKLKGENGLTV-----ATNKDGTVTFTGINTQSGLKGADSTTLNKGDLISIK 963
```

Db 1195 WSVSVANVLQALNGELTISASLTNSANNITGATHDIVIDA-----NLPLRLVD 1244  
QY 964 NPASNEQIQVGADGVKFA--AKVDKGNSS---TGIDGTSRITKDQIGFTG-ANGSLDTRKP 1017  
Db 1245 TVAGDDVI---NSIEHTQALVITGSSGLAAGAAALTVVINSVYTGATVLADGWSVGV 1300  
QY 1018 HLTKOKLKVGEVEITWGINAGKKITNTOSGDIQTQNSDNDVGGVRYDLKTELESKINS 1077  
Db 1301 VADVTPNPACTVNIASVGTNTAG-TTTSISHPVTVDLAAVAIT---INTLST--DDVINA 1354  
QY 1078 AAKTAQNSLHEFSADEOHNHTVSNPYSDTSKTSVITFAGENGITTKYNGKGVVRVG 1137  
Db 1355 A-----EKSDQLQSGTTSVGEAGQTITVI-FGGKSYTTT-----VA 1390  
QY 1138 IDQTKGLTPKLTGVNNGKGVIDSKDG-QNTITGLSTLANVTNDGAGHALSQGLAND 1196  
Db 1391 ADNTWGLTTPAVDVAT-----LPDGAANVQASVSNVAGNSTQ--ATHAYSVDATAP 1439  
QY 1197 TDKTRAASGDVINA-----GFTNLOGN----- 1218  
Db 1440 SVTINTIATDDTLNAAEAGSALTISGTSAEAGQTVTVTLNGVNYSGNVQADGWSVSV 1499  
QY 1219 -GEADVSTYDTDFIDGNATTAKVYDDTSKTSKVYVDVNVNKTIT- 1265  
Db 1500 TGDLASLTASSYTV-----NASVSKARNSASATHNLTVDLAAPPVTVINTVAGDDIINAT 1554  
QY 1266 -----EVTSDKKLG-----VKTTTLTKT-SANGN-----ATKFS-AADGDAVK 1302  
Db 1555 EHQQAIIIGSATGATTGNTVSTGTTTITVTVLNGANGWSIGVPSVISAQAQGVIT 1614  
QY 1303 A-----SLATH-----INTLAGD--IQTA-KGASQASSASYVDADGNK 1339  
Db 1615 AVTDSAGSGTASHTVTVVALGAPVLAINITIAVDIINAAEKAGDLAITGTSNQPA-GTQ 1673  
QY 1340 VIYDSTDKKIYQVND-KGOVDKNKEVAKDLVAAQ-----TPDGTLAQNMVSVINKEQVN 1395  
Db 1674 IIVTLNGQNYTTADAGSNWSVTVPASRSALGEATYTVTAATDADGNSGSASHNVQVN 1733  
QY 1396 DANKKOGIN-----ED--NAFTKGLN-----AAKDTKTNAAVTVG--DLNAVAQT 1438  
Db 1734 TALPGVTINVATDDIILNAAEAGVEQTISGQVTGAAGDTVT-----VTLGGATYATVQA 1789  
QY 1439 PLTEAGDGTGA-KLIGE-TLTIKGGQDITNKLTONNIGVAGTGFVVKLAKDLTNLS 1496  
Db 1790 NLWSVDVPASALQELNGELTISASVNSVNTGN-----GTREITI-----D 1833  
QY 1497 VNAGGTRID-----ERKISFVDANGQAKANTPVLSANGLDLGGKRTSNIGCAAVDDNDVNF 1552  
Db 1834 ANLPGLRVDTVAGDDVNIIE-HGQALVITG--SSSGLAAG-----SNTLTI----- 1878  
QY 1553 KOFNEVAKTVNLLNQNSGASLPEVVTIDAN---GKPIINGTDGKPKAIK-GADGKYHYA 1608  
Db 1879 -----NQCTYVAA-----VLADGTWISVGPVAVDVSAWPAGSVTTAAAGSGTSAG 1921  
QY 1609 NANGV--PVDKDGKPI-----DADKLANIAAHCKPL-----DAGHQVVASLGCN 1651  
Db 1922 NPVSVTHPVTDLSAVAVSINAITADDVINAERGAALTLTSGTSGVAGQTVTVTFGGK 1981  
QY 1652 SDAITLTNIKSTLPQIDTPNTGNAGQAQSLPS--LSAAQOSNAAVSKVDLVNFGNLOT 1709  
Db 1982 TVSATVA-----ANGSWSTSVPAADMAALRDGDASAAQASVSNVNGNSAT 2025  
QY 1710 NIN--QVDFVKAYDTVNFVNG-----TGADITSVSAD-----GTMSNITVN-----TAL 1752  
Db 2026 TTHAYSVDSAPTIVTINTIAGDDIILNAAEAGAAALTTIGSSTAEGQTVTVTLNGTNYTGT 2085  
QY 1753 AATDDGNNVLKADGKFKYKADLLMPNGL--KAKKSAS---DAKTPGLSLVNPCKAG 1807  
Db 2086 VOTDGSWSVPSADSLTSLTASXTVNAVSDSKAGNPVSNHNLTVTSPVPTVINTVAG 2145  
QY 1808 STGDVALNLSKAVFKSKDGTITTTVSSDGGISIQGKDNSSITLSKDG---LNVGKGVIS 1864

Db 2146 D--DVINATEHAQAQIIISGSATGAATGSTVTVTI-CTNTFTTFLVDASGNWSVGPASVVS 2202  
QY 1865 NVGKGTKDDTDAANVQQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSSRNRYTK 1924  
Db 2203 ALANGTVTI-----NASVTDAG-----NSGSATHQVTVN 2232  
QY 1925 AG-----TVLGGKGNNDTEK-----LATGCVQVGVKDKG---NANGDLSNV 1962  
Db 2233 TGLPTITFNAISGNDILNADEKQGPLTISGGSTGLAT-GAQVTVTLNGHNSATTDASGN 2291  
QY 1963 WYTKOKDSKKALLATYNAGQTNILT-----NPAEAIIDRINEOGIRFFHVNQGN 2013  
Db 2292 WTLTVPSDLAAL-----GOANYTVSASATSAAGNTASSOANILLVDSGLPDVTINTVA 2344  
QY 2014 QBPVVOGRN-GIDSSASG-----KHSVAI---GFOAK 2041  
Db 2345 GDDIINAAEAGADQITISGVVTRAAAGDVTVTVLGGNTYTATVQSNLSWSVSVPTADLQAL 2404  
QY 2042 ADGEAAVA-----IGRTOAGNQSIAGDNA-----QATGD---QSTAICTGNVVTGKH 2087  
Db 2405 GNGDLTITASVTNANGTSGSTRDITIDANLPGLRVDTVAGDDIVNSIEHQALVITGS 2464  
QY 2088 SG-----AIGDPSTVKADNSYSGVNNNOFIDATQTDVFCVG----- 2123  
Db 2465 SGLNAGAVLTVTINSVAYSATVQADGWSVG-----IPAANVSAMPAGPLTVEVDQSSA 2519  
QY 2124 -NNITVT-----ESNSVALGNSAIS-----AGTHAGTQAKKSDGTAGTTTATAGATGVKG 2173  
Db 2520 NNPVSVSHPTVLDLTAVAISINTVASDDVINNAEKGNTLTLSGSTSG--IESGQTVTVT- 2576  
QY 2174 FAGQTVAGAVSV-----GASGAERRIQNVAAAGEVATSATSDAVNGSQL 2215  
Db 2577 FGGKTYTASVAANGWSVNVPAADLATLPEGAANVQASVSSASGNSASATHAYSVDASAP 2636  
QY 2216 YKATOGIANATNELDHRHQENKANAGISSAMAMASPOAYIPGRSMVTGGIATHINGQG 2275  
Db 2637 TLTINTIAS-----DDIL-----NAAEAGSPLTISGTSTAE-----TGQTVTVTLNG 2678  
QY 2276 AVAVGLSKLSDNGQVWFKINGSA-----DTQGHVGA 2307  
Db 2679 ATYTG--TVQADGWSVSVPTSALGALNASVTVSATVNDKAGNPGSA 2724  
RESULT 5  
A:14177  
190K surface antigen precursor - Rickettsia rickettsii  
C:Species: Rickettsia rickettsii  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 08-Oct-1999  
C:Accession: A41477  
R:Anderson, B.E.; McDonald, G.A.; Jones, D.C.; Regnery, R.L.  
Infected. Immun. 58, 2760-2769, 1990  
A:Title: A protective protein antigen of Rickettsia rickettsii has tandemly repeated,  
A:Reference number: A41477; MUID:90354033  
A:Accession: A41477  
A:Molecule type: DNA  
A:Residues: 1-2249 <AND>  
A:Cross-references: GB:M31227; NID:g152465; PIDN:AAA26380.1; PID:g152466  
A:Note: The authors translated the codon GAT for residue 430 as Gly, and CAA for resi  
C:Keywords: surface antigen; tandem repeat  
F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>  
Query Match 5.5%; Score 642.5; DB 2; Length 2249;  
Best Local Similarity 23.6%; Pred. No. 5.6e-17;  
Matches 522; Conservative 210; Mismatches 776; Indels 707; Gaps 114;  
QY 261 ALGADTRVDLDYGTALGYSOILANNNNNNKAYVPBGNGSNKSKATGNGLSFSGSST 320  
Db 36 ALGVATGY-----IATNNNAAFSNNVGNMNNNEITAAAGVANGTPAGGPQNNWAFYGGD- 89  
QY 321 IKRKIIINVAGYEDTDVAVNA-----OLKAVENLAKRQITFEKDDNGTGVYKKLGETLTI 375  
Db 90 ---YTVTADAADRIIKAINVAGTTPVGLNITQNTVGSIIITKG---NLLPVTLNAGKSLTL 144



C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
C:Accession: B71704  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, T.; NATURE 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499  
A:Accession: B71704  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2340 <AND>  
A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PID:CAAL4908.1; PID:g386100  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: sca3; RP451

Query Match 4.9%; Score 571.5; DB 2; Length 2340;  
Best Local Similarity 21.7%; Pred. No. 3.1e-14;  
Matches 554; Conservative 295; Mismatches 935; Indels 773; Gaps 133;

Qy 59 ATNGSAVQNNKTAFT-TGNNDNASNEASTAIGSLAKAHANQAIAGGSKPDPRN 117  
Dy 14 ASLCGTLETSN---ATGTTIPNGSVSLNTDAGLVGVFNNGDIIQ-IVNGGREIKISA 69  
Qy 118 QAAQKACSHAKGESIAIGDVLAEADASTAIG---SDDLYLDRNSTNSKYPNGLIST 173  
Dy 70 DKAHAIIGIINTLKELPDFGG---VEVSQNVSIGPLNAGEDL-----NINF--GPLKF 117  
Qy 174 LIQNHVLRQIRDSNGSQYRRRTAAEGHASTAVGAMAYAKGFANAFGRSTAEQNSLA 233  
Dy 118 ISNVVTSI-----ITGVGPKTESNIDFA---GKNATLOINKDL- 152  
Qy 234 VGLFAKAEKGYTIAIGSNAQAINYGALAGADTRVDLDYGTIA---LGYGSOILN-NNNN 289  
Dy 153 -NITTKID-----NTVAGNNGSITFEFS-----GIISNHIIGYTSLLGIVNGC 195  
Qy 290 NKAIVPENGSNLKSSKATNGLFSGSSTIKRKIIINVAGYEDTDVAVNAQLKAVENL 349  
Dy 196 EAKYAPANNITINAK-----INTHNSILTL 225  
Qy 350 AKROI-TEFGDNGT-----GVKKLGETLTKGETQADKL-TDNNNIGVVTDN 398  
Dy 226 CDGNITLLKGNINTEIDGOGILNLAYDLGSSSIITGSDIGNIGSLDITVLLGSATFS 285  
Qy 399 TGLKVKLAKNLSGLTSTKMLTASEKVTGVS--GNNTAELOS----- 439  
Dy 286 TILK---ATNINLKHNTSLNLD-DNIIIVIGNIKGNKNKIDILNFKVHGNTLDNEMIIPAP 341  
Qy 440 ---GGTFTPTTNASTDKTGVYGDGLKFTDNTSNTALEDTTTRITKDKIGFSNK---AGTV 492  
Dy 342 QKTHTLNF--KGNATLNGNTNNILKFSGGHKTMLQGNTKVDNLVLFADSLDSTI 399  
Qy 493 DENKPYLDKDKLVGNSTLNGGLTVN--NTI-----GGSNKQIQVAGDGKIFADVNVV 545  
Dy 400 SVN-CLLTDCTVFNSNVNGTLLINAKNTISAKLLNATKAKIQIN-----ANLTNNH 452  
Qy 546 SNAKFGTTRITEEIGPADA-DGKYDKSPYLDKQLQVGVKITKDSG-----INAG 598  
Dy 453 PSAGDISIRIADNTIYTDARKGNV-----LLNNAKIIIFEGADSMALALNTG 502  
Qy 599 DQKISNVKDATDDTDAVYKOLKVOQDADGALQSFISIRDEKQGQEFFISNLSNGTNTPT 658  
Dy 503 -----VTADRTFTIYNLNGSNDYEIVKIEAK----- 532  
Qy 659 PETTFAGENG-ISISND-----IAKGVKVGIDIPINGLTPKLTGVSKDKGTOLV 709  
Dy 533 -KVITIANOSPGYITQDNTHRLAKELIVEGAGDIIID--DTIFTKLLSINSTGTQITENRT 589  
Qy 710 IEQVAGSNDTKN-----IIRGLSPLPSITNAGVVRITEQGN---TITSDKSKAASIG 761  
Dy 590 LDLAGGNIARFGKGTLLVNGVGTSTTSENNQGIITNSGNTITGVIGTNGELGLKLVNIG 649  
Qy 762 -DILNTGFLNKNNSVGFVSTYNTVDPIDGNATTAKVYDETNTQTSKVTYDVNVNDEKTI 820

Dy 650 ADPVTCSANV---FASVALTPSSVLLIADGVTLTGEVT---THNNTKGVLSLG----- 697  
Qy 821 ELTDGNGTKNGKVTTLTTTNNANGKATNFS-----TTDNDALVNAK 863  
Dy 698 --TGSN-ITGQIGTNSAALEKINICAGASINISNIYAGSVLTDTOTSELTNNDDVVNSN 754  
Qy 864 DIAENLTLAKEIHTTKTADTALQTFVKKDGATDDTETIVGKDGTON-GKTVMTLKLK 922  
Dy 755 IITTAGNNSGKLIETGNGGI-----TGNIGANGAALQEVVF---NGTTNIGGTANS----- 802  
Qy 923 GENGLTVATNKGDTVTFGINTOSGLKAGDSTTLNKG-----LSIKNPASNFOIQVGADG 977  
Dy 803 --QNFTVAHSAANVITGLTT-GALKYKDTGTIIAHGLVGDIDFNNKAG----- 849  
Qy 978 VKFAKVDKGNSSGTGIDTSRITKQIGFTGANGSLDTTKPHLTKDKLVGVEVIT-NTGI 1036  
Dy 850 -KFILGD---GAMIDGSLCN-----GGVAGTLDF-----IGDGNVTQNIKA 887  
Qy 1037 -NAGKKTYNIOSGDIJON---SND-----VTGGRVYDLKTELESKINSAAKTAQNSL 1086  
Dy 888 DNANSISTINIQ-GDNTKNVTIANDIFVDNIHFTNGGILQL----- 927  
Qy 1087 HEFSVADQGHFTVSNPYSYDTSKTSDVITFAGENGITTKVKNKVVRVGDIDOTKGLTT 1146  
Dy 928 -----GGNLTHN-----IDF-GANGGTLEFN----- 948  
Qy 1147 PKLTVGN--NGKIVIDSQDQNTITGLSNTLANVTNDGAGHALSQGLANDTDTK-RAA 1203  
Dy 949 -----GNNTYNLNAIIV--NGQN--GILNAFTNLK-----ASDDTIGTVKII 986  
Qy 1204 SIGDV-LNAGNPLAGCEAVDPVSTYD--TVDFIDGNA--TTAKVTVYDDTSKTSK----- 1253  
Dy 987 NIGQITGPQNTFQIIVNNKNLTLVSSVNSINFGDANSOLILSAPVDQTIKFINNLETGG 1046  
Qy 1254 --VYDVNVNDKTTTEVTSDDKLGKVTTLTKTSANG-----NATKFSAA 1295  
Dy 1047 GIITLDSNGNLTISGNGIKLGSKNELSLNIGKVTVTNDDLIQNIHQLINNNGALF 1106  
Qy 1296 DGDALVKASDIATHLTLAG-----DIQTA-----KGASQAS-- 1327  
Dy 1107 DDQSLTSAKIRKINIGTVAGGATYTLDAINDFNLTSQWVFKHODSILELKNSSNTNDH 1166  
Qy 1328 --SASVYVADGKN--VLYDSTDKKYOVNDKQVDDKKEVAKD--KLVAQAQTPDGPILA 1381  
Dy 1167 TITLTSALDPGNNQFGIILITDNTKLITIDNNGNVAYTLGTANIHMLKQLTFASIDNGAIA 1226  
Qy 1382 QMNVKSVINKBOVDANKKOGINEDNAFIKLENAAKDKTKNAAVT-VGDLNAVAQTP 1440  
Dy 1227 ---LKVGINVENVTNLKIDIELNEVNAV--LFN-----KNTTYTATGNING-----HV 1270  
Qy 1441 TFAGDTGTTAKKLGELTLIKGGQTDYTK-----LTDNNIGVVAGTDGFTVKL---AKDLTN 1493  
Dy 1271 DFOGNAGVI--NLNDDIEIDGVSITSGNVGTLPNGSGKVTGLINNIVMLQAGAGDYSL 1328  
Qy 1494 LNSVNAAGTRIDEKISFVDANGAKANTPVLSANGIDL-----GKRIISNIG--AAVD 1545  
Dy 1329 SASGYSITETOGGNNNLTPAANSHLTTDKNTGGQDLNLVFFINGGVSVSIGANAAGV 1388  
Qy 1546 D---NDAVNFKQFNEVAKTVNNLNQNSGASLP-----FVVTDANGKPIGCTDCKPQKA 1597  
Dy 1389 DIIINAGSVNF-----SNTLKSIGNIVISDGCATQVNNNVTATDISGKNAN----- 1433  
Qy 1598 IKGADGKYHYHANAGVPVDKCKPITDADKLANLAHCKPLDAGHQVVASLGGNSDATLL 1657  
Dy 1434 -----NGTLKLNHTPINITSTLGNNAIG-----TIEVANNDVTITG 1471  
Qy 1658 TNIKSTLPQIDTPNTGNANAGQSLPSLSAAQOSNAASVKDVLNVGNLQTNHNNQVDFV 1717  
Dy 1472 T-----LQANITHFSNATQAATL-TLGAASQ-----VTNITTAGNNIHTLE----- 1512  
Qy 1718 KAYDTVN--FVNGTGADITSVR-SADGTMSNITVN-----TALAATDDDGKVLKAKDG 1768



Db 1372 SVTATNPT-----GNSAT--ATQTIV-----VDTTAPT 1397  
Qy 1019 LTKDLKVE-----VETNTGINAGKKIIRNIQSGDITQNSNDVATGGRYVD 1066  
Db 1398 VTINAIAVDIIINAVEAGSPVAVSGTTGVEDG-----QVVTVTIDGNTYATATGNAWT 1452  
Qy 1067 LKTELESKINSAAKTAQNSLHEFSVADGQNHFTVSNPYSSYDTSKTSVDITFACENGIT 1126  
Db 1453 FNIPIADIANFEATEEVA-----TVSDLAGNPATPATRNIITVDTTAPTVTINAIAVDII 1508  
Qy 1127 TKVNGGVVRVIGIDTKGLTTPKLTGVNNNGK--GVIDSKDCQNTITGLSNT-----LANV 1180  
Db 1509 NAVEAG-SPVAVSGT-----TTGVEDGQVVTVIDGNTYATATGNAWTNIPVADI 1559  
Qy 1181 TN-----DGAGHALSQGLANDTDKTRAASIG-DVLNAGFNLOGNEAVDFVST 1227  
Db 1560 ANFEATEEVAIVTSVSLAGNPATPATRNIITVDTVAPAVTIDSISDDTGAQAN-----DFITN 1615  
Qy 1228 YTVDFIDGNA--TTAKVTYDDT--SKTSKVYDVAVD-NKTIETVSDKKLGVKYKTTIL 1280  
Db 1616 DTLVFP-NGTAEDSTVVVSLDGIIGTGTANGAGETWLDYGTLLADGDVLSVATNP 1674  
Qy 1281 TKTSANGNATKFSAADGDALYKASDIATHLNTLAGD--IQAKGASQASSASYSYVDADGN 1338  
Db 1675 TGNSTATATQT-----IVVDTTAPTVTINAIAVDIIINAVEAGSPVAVSGTTGVEDG 1726  
Qy 1339 KVIYDSTDKKYQVNDKQGVNDKKEVAKDLVAQAOTPDGTLAQMNVKSVINKEQVNDAN 1398  
Db 1727 QVVTVIDGNTYATVTC-----NAWTFNIPVADIANFEATEEVV 1766  
Qy 1399 KKGINEDNAFIKLENAKDTKTNA--AVTVGD-LNAV-AOTPLTFAGDTGTAKKL 1453  
Db 1767 ATVSLAGNPATPATRNIITVDTTAPTVTINAIAVDIIINAVEAGSPVAVSGT--ITGVED 1824  
Qy 1454 GETLIK-GGQTDTKNLTDN-----NIGV-----VAG-----TDGFT 1484  
Db 1825 GOVVTVIDGNTYATVTCNAWTFNIPVADIANFEATEEVVATVSDLAGNPATPATRNI 1884  
Qy 1485 VKLAKDLNLNSVAGGTRIDEKGISFVD-----ANGQAKA-NTPVLSANGLDLGG 1534  
Db 1885 VDTVAPVTTDISIS-----DQTGAQANDFITNDTFLVNGTAEADSTVVVSLDGIETG 1938  
Qy 1535 KRISNIGAAVDNDNAVNFQNEVAKTVNNLNQNSCASLPFVVVTDANGRP----- 1586  
Db 1939 VTANGAGETWLDYGTLLAD-GDYELSTATNPTGNSATATQTIVVDTTAPTVTINAIAV 1997  
Qy 1587 ---INGTDGKPKQAKGA-----DGKYYHANANG-----VPV----- 1615  
Db 1998 DDIINAVEAGSPVAVSGTTGVEDGQVVTVIDGNTYATVTCNAWTFNIPVADIANFEA 2057  
Qy 1616 -----DKDGRPIDA-----DKLANLAHGKP----- 1637  
Db 2058 TEVVATVSDLAGNPATPATRNIITVDTTAPTVTINAIAVDIIINAVEAGSPVAVSGTTG 2117  
Qy 1638 LDAGHQVVASLGGNSDALTLLNIKST--LPQIDTPN-----TGNMAGAOQ 1681  
Db 2118 VEDGQVVTVIDGNTYATVTCNAWTFNIPVADIANFEATEEVVATVSDLAGNPATPATR 2177  
Qy 1682 SLPSLSAAQQAASVSKDLVNGFNLOPNHNOVDFVKAYDTVNFVNGTG-ADITSVRSAD 1740  
Db 2178 NITVDTVAPAVTIDSISD-----DTGAQANDFITNDTFLV-NGTAEDSTVVVSLD 2228  
Qy 1741 G-TWSNITVNTALATDDGNVLKADGKFKYKADDLMPNSGLKAGKSASDAKTPTGLSL 1799  
Db 2229 GIEGTVTANGAGETWLDYGTLL--ADGDY-----ELSVTATNPTGNSATATQT-----I 2277  
Qy 1800 VNPAGKSGTSDAVALNNLSKAVFKSK-----DGTFTTTTVSSDGISIQCKDSSITLSKDG 1855  
Db 2278 VVDITAPTVTINAIAVDIIINAVEAGSPVAVSGTT-----GV-----EDGQVVTVIDG 2327  
Qy 1856 LNVGKVTISVNGK-GTKDTPDAANVOQLNEVRNLLGLGNAGNDNADGNQVNTIADIKDPNS 1914  
Db 2328 NITYTATVTCNAWTFNIPVADIANFEATEEV-----VATVSDLAGNPAT 2370

Qy 1915 GSSSNRTV-----IKAGTVLGGKGNNDTEKLATGCGVQVG-----VD 1950  
Db 2371 PATRNIITVDTTAPTVTINAIAVDIIINAVEAGSPVAVSG-----TTTGVEDGQVVT 2423  
Qy 1951 KQGN-----ANGDLSNVVVTQDKGSKKALLATYNAAGQNTYLTN 1990  
Db 2424 IDGNTYTTATVTCNAWTFNIPVADIANFEATEEVVATVSD-----LAGNPATPATRNI 2477  
Qy 1991 N---PAEAIDRINEQ---GIRFFHVND-----GHOE---PVVOGRNGID-----SSASG 2030  
Db 2478 DTVAPAVTIDSISDDTGAQANDFITNDTFLVNGTAEADSTVVVSLDGIETGTANGAG 2537  
Qy 2031 KHSVAIGFOAKADGEAAVAIGRQTOAGN-----QSIATGDNNAQATGDOSIAI----- 2077  
Db 2538 EWTLDYGTLLADGDYELSVATNPTGNSATATQTIVVDTTAPTVTINAIAVDIIINAVE 2597  
Qy 2078 -GTGNVVTGKHSGA-IGDPSTVKAD-NSYS---VGNNNQF-----IDATQ 2116  
Db 2598 AGSPVAVSGTTGVEDGQVVTVIDGNTYATVTCNAWTFNIPVADIANFEATEEVVATV 2657  
Qy 2117 TDVFG-----VGNNTVTES-----NSVALGSNSAISAGTHAGTQAKKSD 2156  
Db 2658 SDLAGNPATPATRNIITVDTTAPTVTINAIAVDIIINAVEAGSPVAVS-GTTGVE---D 2712  
Qy 2157 GTAGTTTATAG--ATGTVKGFAGQTAAGVAVSGASCAERRIONVAAGEVSATSTDAVNGSQ 2214  
Db 2713 GQVVTVIDGNTYATVTCNAWTFNIPVADIA-----NFEATEEVVATVSD-LAGNP 2763  
Qy 2215 LYKATQGI-----ANATNELD 2230  
Db 2764 ATPATRNIITVDTVAPAVNVELD 2784  
RESULT 8  
C48399  
ABC-type transport protein ydbA.2 - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 19-Nov-1993 #sequence\_revision 16-Oct-1998 #text\_change 08-Oct-1999  
C:Accession: C48399; D64891; H64891  
R:Moszer, I.; Glaser, P.; Danchin, A.  
Biochimie 73, 1361-1374, 1991  
A:Title: Multiple IS insertion sequences near the replication terminus in Escherichia  
A:Reference number: A48399; MUID:92190338  
A:Accession: C48399  
A:Molecule type: DNA  
A:Residues: 464-2020 <MOS>  
A:Cross-references: GB:D85081; NID:g3041754  
A:Experimental source: strain K-12  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIN:88089, NCBI:88090)  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: D64891  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-839, 'LDLPLYFOTSVIT' <BLA1>  
A:Cross-references: GB:AE000237; GB:U00096; NID:g1787665; PIDN:AAC74483.1; PID:g17876  
A:Experimental source: strain K-12, substrain M61655  
A:Accession: H64891  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 'M', 915-2020 <BLA2>  
A:Cross-references: GB:AE000237; GB:U00096; NID:g1787665; PIDN:AAC74487.1; PID:g17876  
A:Experimental source: strain K-12, substrain M61655  
C:Genetics:  
A:Gene: ydbA\_2  
A:Start codon: GTG





[illegible]

QY 1429 VGDINAVAQPLTFAGDTGTTAKKLGELT- 1458  
Db 1529 AKDLNL-----KVGSLTKEQDNKLEKIDIVYVKTENVGEVLVPQVY 1571  
QY 1459 -----IKGOTDTNKLTDNNIGVVAG-----TGFTV 1485  
Db 1572 LAKQTEEVKEQKRGVGTQIRAGIDVKKVDDVNTGTIAGVCLAKNKLKNTGDLISQ 1631  
QY 1486 KLAKDLTNLSVNAAGTRIDEKISFY-----DAN-----QAKANTP 1523  
Db 1632 RLSLKGKKGLESTGVYVDTGATKVRKARIKSEGHILETOKDNVDLTASELKGNTG 1691  
QY 1524 VLSANGDLG-----GKRISNICAADVNN-----DAVNKQFN 1556  
Db 1692 QIRAKDLNLNDIYETSYKRYKELFGKNGGEIGDRVTQTSQAQSVGTDAFSDLHLHSLG 1751  
QY 1557 EVAKTVNNLNQNSGASLPFVYTDANGKPIGTGDKPOKAIKAGDKYYHANAGVPVD 1616  
Db 1752 DVNQTGSLKANRTTGV-----VKGFNTKA--CKDLFHRQIDVTSGTVYSASAG---- 1801  
QY 1617 KDGKPTDADKLANLAHAGKPLDAGHQVVASLGGNSDAITLTNIKSTLPQIDTPNT----- 1672  
Db 1802 -----GGQSAGISLTDQGVETYNKTATAGANAD---VTNFMKRTRETETSLTHRS 1850  
QY 1673 -----GNANAG-----QAQSLPSLSAAQNSAASVKDVLNVGNLQTNHN 1712  
Db 1851 EFNALSGELVYMGKADIGGVDIRNDVEVITPEEIAAEAKAEAEK-----KAEYKEN 1903  
QY 1713 QVDFVKAIDVNFVNGTGADITSVRSADGT---MSNITVNTALAAATDDGCVNLKAKDGK 1769  
Db 1904 EASETAKETEE-----AENDNVAEKDKTKPKFKLUTDEEIAAAFEKGEDEFFAA----- 1953  
QY 1770 FYKADLLMPNGSLKAGKSADAKTPTGLSLVNPAGKSGTGDVAVALNLSKAVFKSKDGT 1829  
Db 1954 -YKA-----REEEDRK-----KGFTLSAEQIEST-----KARDEK 1982  
QY 1830 TTTTVTS--SDGISITQKDNSSITSLKGLNVGKVISNWGKTGTDTAANYQOOLNEVRNLL 1888  
Db 1983 ETTYELKVGVAEAEAHSA---AADAISNKARQIIDTQGLKQDGTVALQEAASDLNL- 2038  
QY 1889 GLGNAGNDNAGN---QVNADIKKDPNSGSSNRTVIKAGTVLGGKGNNDTEKLTGGV 1945  
Db 2039 -----ATGDLACASAKLKFELSTIEK--KSGASDGR-----SILGRLN-----LAARG- 2081  
QY 1946 QVGVDKDGANGD--LSNVVWVKTQKGSKKALLATYNAAGOTNYLTNNPAAEIDRINEQG 2003  
Db 2082 -----GDITLNNV-----ETTENSLSLKARDNVN--- 2106  
QY 2004 IRPFVNDGNOEPVQGRNGIDSSASCKHSAVIGFOAKADGEAAVAIGRQTQAGNQSIAT 2063  
Db 2107 -----VNSGVTE-----QKDESNSQLKV-----TAGASSGCGVMAGGCSAGV 2144  
QY 2064 GDNAQATGDOSIAGTGNVYTGKHSAGI--GDPSTVKADNSYSVNNNOFIDATQTDVFG 2121  
Db 2145 SAGVSGSYNES-----NTESTSHNLSLLKGSRLVRGAKDFNLISSVDVDHLHLDVK 2198  
QY 2122 VGNNTIVTENS--VALGNSAISAGTHAGTQAKKSDGTAGTTTAGA--TCTVKGFAGOT 2178  
Db 2199 DTNVVSKQDSYSRKRGVNVSVSAG-----VGVSAGGARPNVSV----- 2238  
QY 2179 AVGAVSVCASGAERRIQNVAAGEVSATSTDAVNGSOLYKATQGIATNATNLDHRIHONEN 2238  
Db 2239 ---GLGVSAENENSKIYKQOAGISAKRITGEINNLN---TGGYI-----EN 2279  
QY 2239 KANAGISSAMAMAMPQAYIPGRSMVTGGIAT-----HNGQGAVALVGLSLSDNGQWVF 2292  
Db 2280 KGN-----PDELNVAGDITTHLKDHRHKGSGFCGSGVSVSETGTQV 2322  
QY 2293 KINGSADTQGHVGA 2306  
Db 2323 NVNGRVEQKHVEA 2336

RESULT 10  
F81045  
hemagglutinin/hemolysin-related protein NMB1768 [imported] - Neisseria meningitidis (C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: F81045  
R;Tettelein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. ri, H.; Qin, H.; Vanathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M. Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: A81000; MUID:20175755  
A:Accession: F81045  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2514 <TEU>  
A:Cross-references: GB:AE002526; GB:AE002098; NID:g7227015; PIDN:AAF42109.1; PID:g7222 A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB1768

Query Match 4.7%; Score 546; DB 2; Length 2514;  
Best Local Similarity 19.4%; Pred. No. 3.2e-13;  
Matches 524; Conservative 341; Mismatches 971; Indels 860; Gaps 117;

QY 4 IYKVFENKATGTFMAVAECAKSHSGSSSTAGGVSSPVIRLVRATLAILVIGATLNG 63  
Db 5 LYRVFNRKGAHVAVAEITTKREKSCADSD-----SG 37  
QY 64 SAYAQNNSKIAGTGTGNNDNASASNEASIAIGS----LAKAHANQAIAG--GSKPPDRN 117  
Db 38 SARVKS---VPFGTT--HAPVCRSNIFSFLLGFSCLAVGTANIAFADGIADKAAPKT 92  
QY 118 QAA--QKASHAKGKESIAIGGDVLAEGDASIAIGSDLLYLDNRNSNSKYPNGLLSTLI 175  
Db 93 QOATILGTGNGIPQVNTQPTTSAGSVNQVAQFDVNGRGAILLNRSNTQTPQ----LGGWI 149  
QY 176 QNH-----TVLRQIRDSNGSKYRRTAAEGHASTAV----GAMAYAKGFHANAF-G 221  
Db 150 QGNPWLARGEARVVVNGIINSHSSOMNGYIEVGGRAEVVIANPAGIATVNGGGFINASRA 209  
QY 222 TRSTAEGNYSVLAVGLTAKAEKGYTIIAGSNAQA-----INY----GALALGADTRVD 269  
Db 210 TLTTGQPOYQAGDLSGFKIRQGNVVIAGHGLDARDTDFTRILSYHSKIDAPVWQDVRVV 269  
QY 270 LDYGIALCYG---SAILNN-----NNNNNKAYVPEGNSNIKSKATGNGLFSGSST 320  
Db 270 AGQNDVATGNAHSPILLNAAANTSNNTANGTHIP-----LFAIDTG- 312  
QY 321 IKRKIIINVGAGYEDTDVAVNAQLKAVENLAKRQITFRKGDONGTGVKKKLGELTLIKGET 380  
Db 313 -----KLGGMAN-----KITLLISTAEQAGIR-----N 335  
QY 381 QADKLTDNNIGVVTDN---NTGLKVLKAKNLGLETVSTKNLTASEKVVTVGS-----GNN 433  
Db 336 QGQLFASSGNAIDANGRLVNSG-----TMAAANAARDTN-TAEHKVNIRSQGVENS 387  
QY 434 TAELOSGLTFTPTTNASTDVYTGDLKTDNSNTALEDT-TRITKDKIGFSNAGTV 492  
Db 388 TAVSQG-----TOHSOSIQNTGTLSSGEILIHNSGLK 423  
QY 493 DENKPYLDKOKLVGNSTLNNGLTVNNTIGGSNKQIQVGADGKTFADVNVVNSAAKFG 552  
Db 424 NETSGTTEARLAIDTDLNNQ-----KLSQTSQKLHI----- 458  
QY 553 TTRITEEIGFADAGKVDKKSPLYDKKQLQVGVKITKDSGINAGQDKISNVKDAYDDT 612  
Db 459 -----DAQKMDNRGM-----GLQDTAPTASNGSSNOTGNSYNA----- 493  
QY 613 DAVTYKQLQVQDADGALQSFSDERKQGEFTISNLNSNGTNTPTTETITFAGENGISI 672

Db 494 -----SFHSSTTPT---TATGTGATVSI 515  
Qy 673 SN-----DIAGKVKV-----GIDPINGLTTPKLTGSDKDKTQLVIEQV-ASG----- 716  
Db 516 SNITAPTEADGTIRHGHALDNGSGSIANGQDVSAQAQGLNLAG--QIDIHOLNAKGSAPD 573  
Qy 717 -----NDTKNIIRGLSFTPLPSITNAGGVRTTEGNTITSD-----DKSAASIGDIL 764  
Db 574 NHNGTIIIDAVHIQAG-----SLNQNGNITTRQOLEIETDOLDNARHGKLSIAEADLA 627  
Qy 765 NTGPNLKNNSVGVSTYNTWDFIDGNATTAKVTYDNTNOTSKVTYDVNVYDEKTIELTG 824  
Db 628 VSG--SLNQ--GEIATNQQLIHDGQOSTAVI--DWTNGTIOGSRDVAQAQSLSNNG 681  
Qy 825 DNGKTNKIGVK-----TTTLTTTANGKATNFSTTNDALVNAKDI---AENLNTIAK 874  
Db 682 TLAADNKLIALQDDFYVERNIVAGNELSLSTRGSLKNSHTLQAGKRTIRIKANNLDNAAQ 741  
Qy 875 EIHTTKGTADTALQTFYKKGADTDEIITVGKGTQNGKTVNTLKGKENGILVATNKD 934  
Db 742 GNIOGGTDTIGTQ--HNLTNRGLIDGQOTKI-QAQOMN--NIGTGRIYGDNIATAATRLD 797  
Qy 935 GTVFTGINTQSLKAGDSTLTKDGLSTKNPASNEIOVGADGVKFAVDKGNSTGIDG 994  
Db 798 NODENG--TGAIARAENLNL--GIGOLNRENLSLIYSGNDMAVGALDNGQATG--- 849  
Qy 995 TSRTKDOIGFTGA-----NGSLDTRPHLTOKKLKGYEVEITNTGI---NAGCK----- 1041  
Db 850 ---KAQRHNAGATIEAAGKMRGVEKLHNTNEHLKTQLVETGREHIVDYVEAFGRHELL 905  
Qy 1042 -----KITNIQSDITQ-----NSND 1057  
Db 906 REGTOHELGSVYNDESHLRTPGAHAENHMKYDYKVT--QKTQVTTQAPAKIISND 963  
Qy 1058 -AVTGGRYVLDKTELESKINSAAKTAQNSLH-----EFSVADEQGN-HFTVSNPYSSYD 1109  
Db 964 LTIDGKEVNTDSQIIAGNLIIVQTEKGLNEQTFGEKKFSENGKLSHWREKHKRD 1023  
Qy 1110 TSKTSDVITFAGENGITTKVNGGVVR-----VGIDQTKGUTTP----- 1147  
Db 1024 STGHSEQ--NYTLPEITRNIISIGSPAYESHRKALSHHAPSOGTELPQSNGLISPLTSNSF 1082  
Qy 1148 -----KLTGVNNGKGTVIDSK-----DGQNTI 1170  
Db 1083 TPLSSSYLIIPVNGKYLIVETDPRFANVROWLGSYMLDSKLDPNLHRLRGDGYEEQ 1142  
Qy 1171 TGLSNTLANVTNDGAGHALSOGLANDTDKTRA-----ASIGDVLNA----- 1211  
Db 1143 RLINQIAIELT---GHRRLDGYQNDQEFKALMDNGATAARSMNLSVGIALSIAEQVAQL 1198  
Qy 1212 -----GFRNLOGN----- 1218  
Db 1199 TSDIWLVOKEVKLPDGGTQTVLPQVYVRVKNKGIDGKALLSGSNQIINVSGSLKNSG 1258  
Qy 1219 ---GEAVDFVSTYDTPVDIDGNATTAKVTYDDTSTKSVYDVNVNDKTIETVSDKKLGV 1275  
Db 1259 TIAGRNALIINP-DPLDNGIRHIAQKSAVATQDINNIGMISAE-QTLILLNAGNNINS 1316  
Qy 1276 KTTTLTKTSANGNATKFSAADQ-----DALVKASDIATHLNTLAGDI--QTAKGASQ-- 1325  
Db 1317 QSTTASSQNTQGSSTYLDRMAGIYITGKEGVLAQAQAKDINIITAGQISNSEQQTRLQ 1376  
Qy 1326 -----ASSASVYDADGNKVYIDSTDKKYYOVNDKQGV---DKNKEVAKOKLV 1370  
Db 1377 AGRDINLPTVOTSKHQATHPAD-NHVIRGSTNEVSGSIQTKGDVITLLSGNNLNAK---A 1432  
Qy 1371 AQAOPTPDCTLA-----OMNVKSVINKEOVNDANKQGINEDNAFIKGLLENAAKDTKTNA 1425  
Db 1433 AEVSSANGTILAVSAKNDINISAGINTHVDASKHTGRSGGNKLVITDKQAQSHHETAQS 1492  
Qy 1426 AVTGV-----DLNVAQTPPLTFAGDTGTAK-----KLGETLTKGGQTTDNKILTD 1471  
Db 1493 STFEKQGVVLQAGNDANILGSLNVIS---DNGTQIQAGHNVRIGTGT--QSQSEYTHQTQ 1547

Qy 1472 NNIGVVAAGTGTGFTVKLAKDLNLNLNVNAGGTRIDPKGISFVDANGQAQKANTPVLASGLD 1531  
Db 1548 KS-GLMSAGIGFTIG-SKNTNQENQ-----SOSNEHTGSTV---CSLKGDTTIVA----- 1592  
Qy 1532 LGGKRISNIGAARDNDNDAVN--FKQNEVAKTVNNLNQNS-----NSCASLPP--VVTDA 1582  
Db 1593 --GKHYEIGTSTVSPEGNTIYAOSIDIOAHNKLNSNTTQTYEOKGLTVAFSPVDDL 1650  
Qy 1583 NGKPLNGTGDGPRQAKGADGKYIHANAN-GVPVDKDGKPTDADKLANLAHAKPLDAG 1641  
Db 1651 AQAIAVAQSSQKVQOSKNDRVNMAAANAGWAQVOTGK---SAQNLANGTINAKOV--- 1704  
Qy 1642 HOVVASLGGNSDAITLT-----NIKSTLPQIDTPMTNANAGQAQSLPSLSAAQOSNAAS 1696  
Db 1705 -----SISITYGEQONRQTTQVQANQAQASQIQAGGKTTLLIATGAEEOSN--- 1749  
Qy 1697 VKDVLNVCFLNQTNNHNOYDFVKAYDVTNPFVNGTGADIT---SVRSADGTSMNIIVNTAL 1752  
Db 1750 -----INAGSDVAGKAGTILIAD---NDITLQSAE 1777  
Qy 1753 AATDDDGNVLKAKDGKFKYKADDLMPNGSLKAGKSASDAKPTGLSL-VNPNAGRG-STG 1810  
Db 1778 QSNTERGO-----NKSAGWNAGAASVFGQGSLSGVTAGGNVKGSGYNG 1821  
Qy 1811 DAVALLNL-----SKAVFKSKDGTITTTVSSDGLSIOGKDNSSITLSKGLNVLGKGVIS 1864  
Db 1822 DSITHRHSHIGDKGQSLTIQSGGDTTI-----KGAQVRGK-----GVQVNAKNLS 1866  
Qy 1865 NVGKGTDKTDAAVNOOLNEVRNLLGLG-NAGND-----NADGNQ 1902  
Db 1867 IQSVODRETYQSK-QONASAQVTVGYGFSAGGDYSQSKIRADHVSVTEQSGIYAGEDGYQ 1925  
Qy 1903 VNIAIDIKDPNPGSSSNRTVIAKAGTVLGGKGNNDPEKATLQGVQGVKDKGN-----A 1955  
Db 1926 IKV-----GNHTDLKGIIIT-----STQSAEDKGNRFQTATLT 1959  
Qy 1956 NGDLSNVWVKTKQDGSKKALLATYNAAGOT-----NYLTN----- 1990  
Db 1960 HSDIKN---HSQYKGEFSLGASASISGKTLQCGAQNKPONKHLTISVADKNSASSSVGY 2016  
Qy 1991 NPAAEIDRINQGI--REFHVNDGNOEPVQVGRNGIDSSA-----SGKHSVAI-- 2036  
Db 2017 SDSQSSQISITKSGINTRIQTIDENAAQIRLTGKTAQKADIDTNTVTTDTERHSGSLKN 2076  
Qy 2037 -----GFOAKADGEAAVA--ICROTQAGNQSIT-----AIGDNAQATGDQSIATIGTNNV 2083  
Db 2077 TFNKEAVQSELDLQRTVSQDFSKNVQQAQTEINQHLDKLADKAEATAAAEALANGDWE 2136  
Qy 2084 TGRHSGAIGDPSYTKADNSYSVGN-----NNQFIDATQTDVFGVG--NNITVTSNSVAL 2136  
Db 2137 TAKRKAHEAQDAKADN-WQOGKVILNMLASGLAAPTS---GAGIAAATASPAVSVAI 2192  
Qy 2137 GSNSAISAGTHAGTQAKKSDGTAGTTTATAGATGTVKGFAGOTAV-GAVSVGAS----- 2188  
Db 2193 GQHPKDLAQGNANKLTAQETAHVLAHVAVGAVAAGVDNNALAGALSAGSSEAAAPYI 2252  
Qy 2189 -----GAERRIQNVAAGEVSATSDAVNGSOLYKATQCIANATNELDH 2231  
Db 2253 SKWLYGKEKGSDLTAEEKETVTATINVLGTATGAAGVNSATDAAQGSLSNAQSAVEN 2308

RESULT 11

C83339

hypoetical protein PA2462 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: C83339

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa



Qy 1761 VLIRAKDGKFKYKADLL-MPNSGLAKGASDAKPTGLSLVNPAGKSGTGDVALNNLS 1819  
Db 1917 VVTDA--CAVLSASLNSOGGIVSAKAAEIRTS-----LN-NSQKGGISGAGLALVA 1969  
Qy 1820 KAVPKSKDGTGTTTSSVSSDGIISIQKDNSS-----ITLSKDLNUGGKVLISNVGKTKTDIDA 1875  
Db 1970 DLVNSQNCRTAKAIDA-NLKLGDQOGSLVSDTAIALDLRGGLVNVNRAOGLIATPG 2028  
Qy 1876 A-NVOOLNEVRNLLGLGNAGNDNADGNNOVIADIKDPNSGSSSNRTVIKAGTIVLGKG- 1933  
Db 2029 ALLUKQLGV-----DNSGGGEI-----SSDRSPTLAATLSNRGG 2064  
Qy 1934 ---NNDTEKL-----ATGGVQV-----GVDKDGNA 1955  
Db 2065 RVISGDSLTIRIAQALDNSLOGVLSASGGLDVAALVDFDNHSGIVASKGDTHTIGVNRLENE 2124  
Qy 1956 NG-----DLSNVWVYTK-----DQSKKALLATYNAAG-----QTYNLTNNAEALDR 1998  
Db 2125 AGRVVSEGAIDLTAQVSSAKGRIAAKGLQVTVGTTLRQOGGELASQGTTLID-ADSLDN 2183  
Qy 1999 INEQGIRFFHVDGN---QEPVVOGRNGIDSSASKHSVAI-----GFOAKADGEAAV 2048  
Db 2184 RGGGLVS---ADGGVTAEARQIDNRGG-EISSVAKVALAVREQLDNRGKGVIGDSELSL 2238  
Qy 2049 AIGRQTOAGNOSIAIGDNAQATGQDQSTIAIGTCNVVTKGH-----SGAIGDPS----- 2095  
Db 2239 TVQRLNNAQKGLVASRDLGLDGAEL-LENGDGLLSQRLVDVTLSGALDNOGGGALVSE 2297  
Qy 2096 ---TVKADNSVSVGNNO---FIDATQ-----TDVPCVGNNTVTPESN 2132  
Db 2298 ESLTVKADQV-----NNOAGTFSSAGSLVTVSRGELNNOGGLVTVDAGATLNTSGFDSNR 2352  
Qy 2133 SVALGNSAISAGTHAGTQAKKSD--GTAGTTTATAGT----- 2168  
Db 2353 ACLYSKAKAVAIRTCALNNSQKSGTGGTGTVLVAGLVNDNREGRISTRKGTLDANLKGLL 2412  
Qy 2169 ---GTVKGFAGQTA-----VGAVSVGASGABRR---TONVAAGEVSA----- 2204  
Db 2413 QGGGSLVGERGVTLDLNGGTLDNHDLGLVSTPGALLRLQLGVNDNSVGEISSDRAFTL 2472  
Qy 2205 -----TSTDVANGSOLYKATOGIANATNLDHRIHQENKA 2240  
Db 2473 AANTLNNOGGRLLSEALTIRIAKTLNLSLKGQVL--ATYDGLATESQVLDNRAGTIGSKG 2530  
Qy 2241 NAGIS-SAMAMASMPQAYIPGRSMVTGGIATHNGQGV-----AYGLSKLSDNGQWY 2291  
Db 2531 DARISVTSILDNAEQSLVSEGRLELVADQVSNQGRIAARGVLEAAVG-TLLQOQGELV 2589  
Qy 2292 FKINGSAD 2299  
Db 2590 SQ--GSLD 2595  
RESULT 12  
T03099  
N:Alternate names: apomucin  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 02-Jun-2000  
C:Accession: T03099; A40009; A28528; B29789  
R:Ekhardt, A.E.; Timpte, C.S.; DeLuca, A.W.; Hill, R.L.  
J. Biol. Chem. 272, 33204-33210, 1997  
A:Title: The complete cDNA sequence and structural polymorphism of the polypeptide chain  
A:Reference number: 214839; MUID:98070526  
A:Accession: T03099  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-13288 <EC>  
A:Cross-references: EMBL:AF005273; NID:g2581863; PIDN:AAC62527.1; PID:g2581864  
R:Ekhardt, A.E.; Timpte, C.S.; Abernethy, J.L.; Zhao, Y.; Hill, R.L.  
J. Biol. Chem. 266, 9678-9686, 1991  
A:Title: Porcine submaxillary mucin contains a cysteine-rich, carboxyl-terminal domain in

A:Reference number: A40009; MUID:91236743  
A:Accession: A40009  
A:Molecule type: mRNA  
A:Residues: 12139-12167, 'T', 12169-13288 <EC3>  
A:Cross-references: GB:M61883; NID:9454837; PIDN:AAA30998.1; PID:g164374  
R:Timpte, C.S.; Ekhardt, A.E.; Abernethy, J.L.; Hill, R.L.  
J. Biol. Chem. 263, 1081-1088, 1988  
A:Title: Porcine submaxillary gland apomucin contains tandemly repeated, identical se  
A:Reference number: A28528; MUID:88087170  
A:Accession: A28528  
A:Molecule type: mRNA  
A:Residues: 12139-12167, 'T', 12169-12641 <TIM>  
A:Cross-references: GB:M21174; GB:J03512; NID:gl64321; PIDN:AAA30990.1; PID:g552360  
R:Ekhardt, A.E.; Timpte, C.S.; Abernethy, J.L.; Toumadje, A.; Johnson Jr., W.C.; Hill  
J. Biol. Chem. 262, 11339-11344, 1987  
A:Title: Structural properties of porcine submaxillary gland apomucin.  
A:Reference number: A92606; MUID:87280230  
A:Accession: B29789  
A:Molecule type: protein  
A:Residues: 1572-1607 <EC2>  
C:Superfamily: pig submaxillary mucin  
C:Keywords: tandem repeat

Query Match 4.5%; Score 520.5; DB 2; Length 13288;  
Best Local Similarity 20.9%; Pred. No. 2.Be-11;  
Matches 547; Conservative 316; Mismatches 1143; Indels 617; Gaps 114;

Qy 8 IFNATCTFMVAACAKSHSGSSSTAGOVGSSPVIRLTRVATLILVIGATLNGSAYA 67  
Db 1492 IYLOSTGISRSTEE---SGSTGISVIGSGKTSITLITVSARS---SVAGSGTTGTVSG 1545  
Qy 68 QNNSKIAFGTTGNNDNASNEASIAIGSLAKAHANQAIAGGSKPDPR-NQAAQKAGS 126  
Db 1546 -----ASGSGTSSGSGSPGATRASIG-----QKETRRISVAGSGGAPAVSSGAQAAGT 1593  
Qy 127 HAKG-----KESIAIGGDVLAEGDASIAIGDDLYLDRNSTNSKYPNGLLS 172  
Db 1594 SGAGPGTTASSVGTETARPASVAGSG---TTGTVSGASGS-----TGSSSGSPGATGA 1643  
Qy 173 TLIONHTVLRIDRSNSOKYRRTAAEGHASTAVGAMAYAKGHFANAFGTSTAB----- 227  
Db 1644 SIGQPETSRISVAGSSGA-----PAVSSGASQAAGTSCAGPGTTASSVGTETARPASVAG 1698  
Qy 228 ---GNVSLAYGLTAKAE-----KGYTTAIGSNQAQAINYGALGADTRV 268  
Db 1699 SGTGTCTVSGASGTCSSGSPCATGASIGQPETSRISVAGSGGAPAVSSGA-SQAAGTS- 1756  
Qy 269 DLDYCIALGYCSOILNNNNNNKAYVPEGNSNKKSKATGNGLFSIGSSTIKRKIINV 328  
Db 1757 -----GAGPGTTASSVGTETARPASV-AGSGTGTCTVSGASG---STGSSSGSPGATGA 1805  
Qy 329 GAGYEDTDAVNVV---QLKAVENLAKROITEKGDNDNGTCVKKKLGETLITIRKGETQADKL 385  
Db 1806 SIGQPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGT-TASSVGT-----ETARPVS 1858  
Qy 386 TDNNNIGV-----TDNNTGLKVLAKNLGLETSTVSTKNLTASEKVTVSGNNNTAELOQS 439  
Db 1859 AGSGTTGTCTVSGASGTCSSGSPCATGASIGQPETSRISVAGSGGAPAVSSGAQAAGTS 1918  
Qy 440 G-----GLTFT---PTTNASTDKTVYTGDKLFTDN-----SNTALEDTTTRIT 479  
Db 1919 GAGPGTTASSVGTETARPASVAGSGTTGTCTVSGASGTCSSGSPCATGASIGQPETSRIS 1978  
Qy 480 -----KDKIGFSNKAQTVDENKPYLDKDKLKVGNSTLNNGL--TVNNTIGSNKQI 529  
Db 1979 VAGSSGAPAVSSGASQAAGTSCAGP-----GTTASSVGTETARPASVAGSGTTG 2027  
Qy 530 QV-GADGKIKFADVNVVSNAAKFGTTRITEBIGPADADGKVDKSPYLDKQLQVGK 588  
Db 2028 TVSGASG-----STGSSSGSPGATGASIG-----QPETSRSIS 2059







Db 1768 ICGTLADGSGNWSFTCTPCTPLANGTVVNAVAQDPAGNMSGPASTVDTVPATVPIPNMSG 1827  
QY 1760 NVL-----IKAK-----DGFKYKADLMPNGSLKAGKSASDAKPTGLSLVNPNGRKGSTG 1810  
Db 1828 SVITGTAEGVAKVILTDG-----NGNPGTETADSGNWTTP-GTPLAN-----GTVI 1875  
QY 1811 DAVALNNLSKAVPKSGDGTITTV-----SSDG-----ISIQCK 1844  
Db 1876 NAVAEDAAGNA-----SGPASTTVDSVAPSAPLISISADGALLTGAEPNSQVRIVVNGD 1930  
QY 1845 DNSSITLSKDL-----NVGKVISN----- 1865  
Db 1931 TANPIITVVDGAGNFSLPFAPPLITGELIAGVADAGNVSGPATINADLPAPTISVPE 1990  
QY 1866 -----VGRGK-TDIDAANVQQLNEVRNLLGLGNAG-----NDNADGNQV 1903  
Db 1991 AADTWINAAEIGDGIQDVTVRTMQGVQVTVKFAGONGYEAESHVTLTAGDIAAGNLT 2050  
QY 1904 -----NTADIKKDNSSGSSN-----RTVIKAGTVLGGKN-----ND 1936  
Db 2051 LTLTPPGMGFPFEGASTVTADINGGTASTPVPTTIDTTPATPVLSLVGNILTISAEPG 2110  
QY 1937 TEKLAT-----CGV-----QVGVKDKGNANGDLSNVVTKRDKGSKALLATYNAAGQNYLT 1989  
Db 2111 TELTVTVDVGGVATATVTVTADNSGLAS-----LNLTLTDLDIDFSWDQLL-----NA--QVSVVG 2162  
QY 1990 NPAAEADIRNEGIGIRFFHVNDGNQPVVQGRNGIDISSAGSKHSVAIGFOAKADGRAAVA 2049  
Db 2163 RDPAG-----NPSNTASIGVGTSEIQPVITIGNEGLDVLNPLNP-RFGSGTTEPSSVV 2216  
QY 2050 IGRQTOAGNOSTAIGNOAQATQOSTAIGNVNVTKHSGAIGDPSTVKADNSYSVGNNN 2109  
Db 2217 IRVITPALNVEL-LPIQADSSGNSFNLISLTILT-----QLGLN- 2255  
QY 2110 QFIDATQTDVFGVN-----NITVTESNVALGNSAISA-----GTHAGTQAKKSDGTA 2159  
Db 2256 -----ITDLNLGSLFSEMLVSTDSN-----GNDSAAYGTLTPNGLSLNIGQIDVNGTS 2305  
QY 2160 GTTTTAGATG-----VRGFAQATGAVGAVSGASGAERRIQNVAAE-----VSATSTDAV-- 2210  
Db 2306 GDDVLSGANGSSHERINGGDSLDLIFNVGTG-----DHSVAGNGNDTIQITATDFVSI 2357  
QY 2211 -----NGSOLYKATQGIATNATNLDHRIHONENKANAGISSAMAMAMPOAYI 2258  
Db 2358 DGGAGFDTLVLANGIDLDYNAVGVGTLSNL--ERI--DLGKGDGSGVLTLTAAEVDAITD 2413  
QY 2259 PGRSM-VTG-GIATHNGOGAVAVCLSKLSLSDNGQWVFKEING 2296  
Db 2414 ANNTLQITGENNDTLNVVGAVENTGTQL-----ING 2444

RESULT 14  
E85822  
Probable invasin 23135 [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: E85822  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: E85822  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2660 <STO>  
A:Cross-references: GB:AE005174; NID:gl2516151; PIDN:AA057041.1; GSPDB:GN00145; UWGP:231  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: 23135

Query Match 4.3%; Score 504.5; DB 2; Length 2660;  
Best Local Similarity 20.2%; Pred. No. 1.13e-11;  
Matches 550; Conservative 329; Mismatches 979; Indels 869; Gaps 138;  
QY 83 NASASNEASTA-----IGSLAKAHANQAIAGGSKPPRNOAANOKAG---SHAKGKESI 134  
Db 132 NSSGNLEQIASTSQIGSLAEDMS-----EQANMARGWASSOASGAMT- 178  
QY 135 AIGGDVLAE-GDASIAIGSDDLYLDRNSTNSKY-----PGLLSTLIQNHITVLR-- 182  
Db 179 ----DMLSRGRTARITLGVDE--DFSLKNSQDFLHPWYETPDNL--FQSHTLHRTD 228  
QY 183 ---QIRDSNGSQYRRTAEG-----HASTAVGA-----MAYAKGHFANAFG 221  
Db 229 ERTQINNGLGRHFTPTWMSGINFFFDHLSRHSRAGIAGYWRDYLKLSNGYLRLTN 288  
QY 222 TRSTAE--GNVSL--AVGLTAKAE-----KGYTIAI-----GSSNAQ 253  
Db 289 WRSAPELDNDYEAPANGWDVRAEGLPAPWPHLGGKLVLYEQQYCYDEVALFDKDDROSNPH 348  
QY 254 AI-----NYGALALGA-----DTRVOLDY-----GIAL----- 276  
Db 349 AITAGLNYTFFPLMTFSAEQORQKQGENDTIFAVDFTWQPSAMQKOLDPNEVDARSLA 408  
QY 277 GYGSQILNNNN-----NNNKAYV-----PEGNSNIKSKKATGNGL--FSIGSST 320  
Db 409 GSRFDLVDRNNNIVLEYRKKEVLRLTLDPTVGKSGEVKSLVSSLOTKYALKGYNVEATA 468  
QY 321 IKR--KIINVAGY-----EDTDVNVVQOLKAVENLAKRQ-----I 354  
Db 469 LEAAGKVVTGDKILVTLPAYRFTSTPDTNDTWPTEVTAEDVKGNFSNREQSMVVVOAP 528  
QY 355 TFRGDDNGTGKVKKGLGETLTKGGETO-----ADKLTDDNNIGVWTDNNTGLKVKLAKNL 409  
Db 529 TLSQKDSVSLS---SOTLSADSHSTATLTFIAHDAAGNPVIGLV-----LSTRH 575  
QY 410 SGLTSTVKMLTASEKVTYVSGNNNAEL-----OSGGLTFTPTTNASTDKTVYTGDLK-- 463  
Db 576 EGVO-----DITLSMDKNDGDSVYQITLTGAMSGTLTLPQLN-----GVDAAKAP 622  
QY 464 ----FTDNSNTALEDITRITKDIGFSNKAGTV--DEN-KPYLDK-----DKL 504  
Db 623 AVVNIIVSSSRTHSSITKIDKRYLSCNPIEVTVELRDENDKPKVEKQKQOLNTAVSIDNV 682  
QY 505 KVG-----NSTLNNGLTVN-----NT 521  
Db 683 KPGVTTDKETADGVYKATYATYKSGSLTAKLLMQWNEDLHTAGFIIDANPQSAKIAT 742  
QY 522 IGSNKQIQVAGDGIKPADVNVNVSNAAKFGTTRITEEIGFADADGV----- 570  
Db 743 LSASNNGVLANENAAVTSVNV-----ADEGSNPINDHTVTFVAVLSGATSFNNTAKT 797  
QY 571 -----DKKSPYLDKQLOVGVKIKTKDSGIN-----AGDKISNVKDATDDTDAV 615  
Db 798 DVNGLATFDUKS---SKQ-EDNTVEVTLENGKVKQTLIVSVFGDSSSTAQVD----- 843  
QY 616 TYKQLKQVQODA-DGALQSFISIDEKQOETISNLSNGTNPNTFTTIFAGENGISIN 674  
Db 844 LQSKNEVADGDSATMTATVRDAG-----NLL-----NDVKVTF 880  
QY 675 DIAKGVKVGIDPIN---GLTTPKLVGSKDKGTQVLVIEQVAGSNDPTKNI----- 723  
Db 881 NVNSAAAKLSQTEVNSHDGIGATATLT--SLKNG-DYTVTASVSSGSAQOQVFIGDQST 937  
QY 724 RGLSPITLPS---ITNAGGVRTT-----EOGNTITSDEKSKAASIGDILNTGFNLKN-- 772  
Db 938 AALTLSVPISGDIIVTNTAPLHMTATLQDKNGNPL---KDKETITFVPNDVASRFSINSNG 994  
QY 773 ----NNSNVGFVSTYNTVDFRIDGNATTAKVYDETNQTSKVTVYDVNVVDKIETLTGNGK 828  
Db 995 KGMTDSNGTAISLTGTL--AGTHMITARLANSNVSDTQPTFVADKDRVVVLOTSKAE 1052  
QY 829 TNKIGVKTTLTTTNNANGKATNFSTTDNDALVNAKO-----IAENLNTLAKEIHTTKGTAD 884

Db 1053 IINGVDETLTAT-----VKDPEDNVVKNLSVYFR-----TSPAD 1088  
Qy 885 TALQTFKVKDGADDEITVCKDGTQNGKIVNTLKLKCNGLVATNKDGTVPFGINTQ 944  
Db 1089 TQL-----SLNARTN-----ENGIAEVLTK-GTV-LGVHTA 1118  
Qy 945 SGLKAGDSTTLNKDG--LSIKNPASNEQIQV-----GADGVKFAKVDKGNSTSGI 992  
Db 1119 EAILNG-----NRDTKIVNIAPDASNAQVTLNIPAAQVVTNNSDVQLTATVQDPSHPV 1174  
Qy 993 DG-TSRIKDKQ---IGFTGANGSLDTTKPHLTCKLKYGEVEITNTGINAGGKKITNIQS 1048  
Db 1175 AGITVFTMPQDVAAEFLENNGIAITQAN-----GEAHVTLKGGKAGCTHTVATL 1226  
Qy 1049 GDITONSNDATGGGVYDLKTELESKINSAAKTAONSLEHFEVSADQEGNHFTVSNPYSSY 1108  
Db 1227 NN---NASDAQPVTFAVDKDSAVVVLQTSKAEIINGVDETTLT-----ATVKDP---F 1274  
Qy 1109 DTSKTSDVITEA-----GENGITTKVNGGVVRVGDIDQTKGLTTPKLTLYGN 1153  
Db 1275 DNAVKDLQVTFSTNPADPQLSQSKSNTNDSGVAEVTFKGTV-LGV-HTAEATLPN---GN 1329  
Qy 1154 NNGK--GVIDSKGQNT-----ITGLSNT--LANVTDNGAGHALSQGLANDTDKTR 1201  
Db 1330 NDTKIVNIAPDASNAQVTLNIPAAQVVTNNSDVQLTATVQDPSHPVAGITVNTFMPQD 1389  
Qy 1202 AASIGDVLNAGFNLOGNGEAVDEVSTYDVFIDGNA---TTAKVTYDDTSKTSKVVYD 1257  
Db 1390 VA-----ANFTLENNGIATQANGEAHVTLKGGKAGCTHTVATLSNNNTSDSQPVTF- 1441  
Qy 1258 VNVDNKTI-----EVTSDKKLGVKTTTLTKT-----SANGNATKFSAD----- 1296  
Db 1442 --VADTKSALVQLQTSKNEITGN---GVDSATLTATVKQDFQNEVNLPVTFSTASSGLT 1496  
Qy 1297 ---GDAVKASDIATHLNTLAGDIOFAGKASQASSASVVDAGKNVLYDSTDKKYQVN 1353  
Db 1497 LTPGESNTNESGIAQ--ATLAG---VAFGEQTVTASLANNGASDNKTVHFTGD----- 1544  
Qy 1354 DRGQVDKNKEVAKDLVAQAQTPDGTLA---QMNVKYSINKVQVNDANKKQGINEDNAP- 1409  
Db 1545 -----TAAAKIIEITPVPDSIIAGTPQNSSGVITATVV-----DNGGFP 1584  
Qy 1410 IKGL-----ENAKDTKYKNAATVGDLDNVAQAQTPLTAGDTGTTAKKLGETLTIKGQQT 1464  
Db 1585 VRGVTVNETSNAATAEMTNGCOAVTNEOGXATVYTTNRSSTIESGARDPTVEASLENGSS 1644  
Qy 1465 DTNKLTDNNIGVYAGTDGFTV-----KLAKDLTNL-----NSVNAGGTRIDEXGISFVD 1513  
Db 1645 TLS--TSINVNADASTAHLTLQALFDPVSAGDTTNLYEVKDNKGNGVPQOEVTLSVSP 1702  
Qy 1514 ANGOAKANTPVLSAN-----GLDLGCKRISNIGAADVDDDAVNFQKOFNEVAKTVNNLN 1566  
Db 1703 SEGVTPSNNAITVTHDGNFYASFETAXRAGVYQVATLENXDSMQ-----QTVTVVP 1754  
Qy 1567 NGSNGASLPVFTVANGKPI--NGTDGKPKQA--IKGADGKYYHANANG-----VPVD-- 1616  
Db 1755 NVANAEISLA-----ASKDPVTANNNDLTTLTATVADTEG-----NAIANSEVFTTLPDVR 1806  
Qy 1617 -----KDCKITDADKLANLAHAKPLDAGHQVVASL--GGNSDAITLTNIKSTLPQDIT 1669  
Db 1807 ANFTLGDGKVVYDTDEGKAKVTLKGTKAGA--HTVVTASMAGGKSEQLVNFIADTL----- 1860  
Qy 1670 PNTGNANQAQSLPSLSAAQGSNAASVKDVLNVFNLTQTNHNVQDFVKAYDTVNFVNG- 1728  
Db 1861 --TAQVN-----LNVTEDNFIA-----NNVGMTRLQATVTDGNGN 1893  
Qy 1729 --TGADITSVRSADGTMSTNITVNTALATDDGN---VLIKAKDGKFKYKADDLMPNGSLK 1783  
Db 1894 PLANEAVFTTLPADVSASTFLQGGGSAITDINGKAEVTLGSKGTYPVTVSVNNYV--- 1950  
Qy 1784 AGKSADAKTPTGLSLNPNAGKGTGDGAVALLNLSKAVFKSKDGTITTTVSSD----- 1837

Db 1951 ----VSDTKQVTLTA----DAG----TAKLASLTSVYSFVSTTEGATMTASVTDANGNPV 1999  
Qy 1838 -GISIOCKNDSITLSKDLGNVGK-----VISNVGKGTDTDAANVQQLNEVRNLLGLG 1891  
Db 2000 EGIAKNFR-GTSVTLSTSVETDGRFAEILVTSFVGLKTVSASLADKPTFV-----IS 2053  
Qy 1892 NAGNDNAGQOVNIADIKKDPNCGSSSNRTVIKAG-----T 1927  
Db 2054 RLLNAKADINSATLTSLEIPQGVMAQDVAVKARVNDQFGNPILNESVTFSAEPPEHMT 2113  
Qy 1928 VLGGKGNNDTEKLATGGVQGVCDKDN-----ANGDLSNVVVKTKD-----G 1970  
Db 2114 ISQNVSTDTHGIAE--VTMTPERNSSYMKASLANGS-----SYEKDLVVIDQKLTLS 2165  
Qy 1971 SKKALLATYNAAGOTNVLTNPAEADIRNEQGIKF-----PHVNDGNOEPV 2017  
Db 2166 ASSPLIGVNSPTGATLTATLTSACTP-VGQVINFSTVPPEGATLSGGKVRTNSSGOAPV 2224  
Qy 2018 VQGRNGIDSSAGSKHSVAIGFQAKADGEAAVAIGRQTAGNQSTAIGDQAQNTGDSIAI 2077  
Db 2225 V-----LTSNKGVTYTVTASFH-----NGVTIQOTIVK 2253  
Qy 2078 GTGNVTVCKHSGAIDGDPSTYKADNSYSVGNNNQFIDATOTDVEGVGN---NITV---TES 2131  
Db 2254 VTGNSSTAHVASFADPSTTAATNS-----DLSTLKAIVED--GSGNLEGLTVYFALKS 2306  
Qy 2132 NSVALGNSAISAGTHAGTQAKKSDGTAGTTTACATGTVKGFAGQTAAGVAVSVG---AS 2188  
Db 2307 GSATLTSLTAVT-----DQNGIATTSVRCATGTSVTSVAVTTAGGMQTVDTILVA 2356  
Qy 2189 GAERRIQVAAGEVSATSTDAVNGSOLYKATQGTANATNELDRIHONENKANAGISSAM 2248  
Db 2357 GPADASOSVLKNNRSSLGKDFDTSALHLVL-----HDSGNPIKVSLEPEVQ 2405  
Qy 2249 AVASNPQAYIPG-----RSWVTC---GIAT-----HNGQCAVAVGLSKLSD 2286  
Db 2406 SCTNAPYQVVSADIDSKNFSGEYKATVTGGEGEATLIPVLNGVHQAGLSTTIQFTRAED 2465  
Qy 2287 N-GOMVFKING-----SADTQGHVGA 2306  
Db 2466 KIMSCVTVLVNGANLPTTTFPSQGTGA 2492  
RESULT 15  
Hemagglutinin/hemolysin-related protein NMB0493 [Imported] - Neisseria meningitidis (C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: H81193  
R:Tetelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M. Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: A81000; MUID:20175755  
A:Accession: H81193  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2703 <FET>  
A:Cross-references: GB:AB002405; GB:AB002098; NID:g7225708; PIDN:AAF40927.1; PID:g722 A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB0493

Query Match 4.3%; Score 504.5; DB 2; Length 2703;  
Best Local Similarity 20.4%; Pred. No. 1.4e-11;  
Matches 557; Conservative 337; Mismatches 1044; Indels 791; Gaps 136;  
Qy 1 MNH-TYKVFINKATGTFMAVACAKSHSGSSST-----AGQVGSFVTRL 46  
Db 1 MNRTLYKVFVKNHRNCMIATAVENAK-REGKNTADTQAVGILPNDIAGFAGTHISVISF 59



```
Db 1966 GSTTHRHVGS-----TGKTTIRSGDITLKGVLIGKGIQADTRLNHLIE 2012
QY 1870 -TKDTDAANVQQLN-EVRNLLGLGNAGN-----DNA-----DGNQVNI 1905
Db 2013 SVQDTETYQSKQNGNVQTVGYGFSASGSYRQSKVKADHASVTCQSGIYAGEDGYQIKV 2072
QY 1906 ADIKDPNSGSSNRTVIKAGTVLGG-----KGNN--DTEKLATGGVQVGVKDGDNANG- 1957
Db 2073 RD-----NTDLKGGIITSSQSAEDKGNLFQTATLTASDIQNHRSRYEGRSFGI 2120
QY 1958 ----DLSNVW--VKYQKDG---SKKALLATYNAAGOTNYLTNNPAFAIDRINEQGIREFH 2008
Db 2121 GGSFDLNGGWDGTVTDKQGRPTDRISPAAGYSGDSDSKNSTPRSG-----VNTHNI---H 2172
QY 2009 VNDGNQEPVVOGR-----NGIDSSASGKHSVAIGFOAKADGEAAVA--IGRQTQA 2056
Db 2173 LTDEAQLARTGRTAKETEARIYTGIDTETADQHS---GHLKNSFDKDAVAKEINLQREV 2229
QY 2057 GNQSTAIQDQAQATGQSTAIQGTGNVVTCKHSGAI-----GDPSTVKADNSYSVG 2106
Db 2230 TRE---FGRNA-AQAAVAADVADKLNTQSYERYQEARTLLEAELONTDSEAEKAFRASLG 2285
QY 2107 NNNQFIDATQT--DVF--GVCNNITVTESNSVALGSNSATSAGTHAGTQAKKSDGTAGT 2161
Db 2286 QVNAVLAENQSRVDTWKEGGIGRSILHGAAGCLTTGSLGGILAGCGTSLAAPYLDKAAEN 2345
QY 2162 TTTAGATGTVKGFAGOTAVGAVSVG-ASGAERRIONVAAGEYSATSTDAVNGSQLYKATQ 2220
Db 2346 LGPAGK-----AAVNALGGAAGIYATGG-----SGGAVVGANVD-WNNRQLHPKEM 2390
QY 2221 GIANATNE-LDHRHONENKANAGISSAM 2248
Db 2391 ALADKYAEALKREVEKREGRKISSQEAAM 2419
```

Search completed: September 13, 2001, 12:42:28  
Job time: 32814 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:53:37 ; Search time 47.77 Seconds  
(without alignments)  
1659.351 Million cell updates/sec

Title: US-09-361-619-11  
Perfect score: 11694  
Sequence: 1 MNHLYKVFKNKATGTFMAVA.....NGSADTQGHVGAAGAGFHF 2314

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : 'SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	642.5	5.5	2249	1 190K_RICRI	P15921 rickettsia
2	556.5	4.8	2003	1 YDAB_ECOLI	P33666 escherichia
3	492.5	4.2	1656	1 OMPB_RICJA	O08653 r outer mem
4	450.5	3.9	1577	1 HLYA_PROMI	P16466 proteus mir
5	448	3.8	3591	1 FIAB_BORPE	P12255 bordetella
6	445.5	3.8	1643	1 OMPB_RICPR	Q53020 r outer mem
7	444.5	3.8	1654	1 OMPB_RICRI	Q53047 r outer mem
8	438.5	3.7	1645	1 OMPB_RICRY	P65989 r outer mem
9	407.5	3.5	1902	1 P2P_LACPA	Q02470 lactobacilli
10	406	3.5	1848	1 CBPA_CLOCL	P38058 clostridium
11	404	3.5	1608	1 HLYA_SERMA	P15320 serratia ma
12	403	3.4	2334	1 WAPA_BACSU	Q07833 bacillus su
13	392.5	3.4	1902	1 P1P_LACLC	P16271 lactococcus
14	392.5	3.4	1902	1 P2P_LACLA	P15293 lactococcus
15	379	3.2	1183	1 CNA_STAAU	Q53654 staphylococ
16	379	3.2	1569	1 YFJA_ECOLI	P52143 escherichia
17	379	3.2	1902	1 P3P_LACLC	P15292 lactococcus
18	376	3.2	1325	1 YDEK_ECOLI	P32051 escherichia
19	362.5	3.1	1300	1 120K_RICRI	P14914 rickettsia
20	356	3.0	1829	1 FRPC_NEIMC	P55127 nelisseria m
21	349	3.0	1567	1 ICEN_YANCT	P18127 xanthomonas
22	336.5	2.9	1694	1 IGA0_HAEIN	P44969 haemophilus
23	336	2.9	2628	1 HAGA_FORGI	Q51845 porphyromon
24	335	2.9	1286	1 AIDA_ECOLI	Q03155 escherichia
25	333	2.8	1829	1 FRPC_NEIMB	Q91yv5 nelisseria m
26	331.5	2.8	1702	1 IGA2_HAEIN	P45384 haemophilus
27	320.5	2.7	1025	1 SLAP_CAUCR	P35828 caulobacter
28	311.5	2.7	1176	1 SLAP_BACSH	P38537 bacillus sp
29	311	2.7	1228	1 SLAP_BACST	P35825 bacillus st
30	310.5	2.7	1597	1 GTF1_STRDO	P11001 streptococc
31	309.5	2.6	1039	1 AG43_ECOLI	P39180 escherichia
32	309	2.6	2491	1 TALA_DICDI	P54633 dictyosteli
33	308.5	2.6	1849	1 IGA4_HAEIN	P45386 haemophilus

RESULT 1

190K_RICRI	1	TOXA_CLODI	2710	1	P16154 clostridium
AC	P15921	1 YDAB_ECOLI	918	1	P34487 caenorhabdi
DT	01-APR-1990 (Rel. 14, Created)	1 ICMA_PANAN	2.6	1	P20469 pantoea ana
DT	01-APR-1990 (Rel. 14, Last sequence update)	1 GTF2_STRDO	2.6	1	P27470 streptococc
DT	01-OCT-1994 (Rel. 30, Last annotation update)	1 APU_THETU	2.5	1	P38536 t amylpull
DE	190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN)	1 PGCY_CHICK	2.5	1	Q90953 gallus gall
OS	Rickettsia rickettsii	1 ICEN_ERWHE	2.4	1	P16239 erwinia her
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;	1 YD86_SCHPO	2.4	1	Q10411 schizosacch
OX	Rickettsiaceae; Rickettsiidae; Rickettsia	1 SLAP_CAMEE	2.4	1	P35827 campylobact
RN	[1]	1 MSB2_YEAST	2.4	1	P32334 saccharomyc
RP	SEQUENCE FROM N.A.	1 GTFD_STRMU	2.4	1	P49331 streptococc
RC	STRAIN-R;	1 ICBV_PSEX	2.3	1	O33479 pseudomonas
RX	MEDLINE=90354033; PubMed=2117568;				
RA	Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;				
RT	"A protective protein antigen of Rickettsia rickettsii has tandemly				
RL	repeated, near-identical sequences.";				
CC	Infected. Immun. 58:2760-2769(1990).				
CC	-!- FUNCTION: ELICITS PROTECTIVE IMMUNITY.				
CC	-!- PTM: GLYCOSYLATED (POSSIBLE).				
CC	-!- DISEASE: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PATHOGENESIS OF				
CC	AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; M31227; AAA26380.1; -				
DR	PIR; A41477; A41477.				
KW	Antigen; Repeat; Signal; S-layer; Glycoprotein.				
FT	SIGNAL	1	28	POTENTIAL.	
FT	CHAIN	29	2249	190 KDA ANTIGEN.	
FT	DOMAIN	212	1180	13 X APPROXIMATE TANDEM REPEATS.	
FT	REPEAT	212	286	A (TYPE I).	
FT	REPEAT	287	358	B (TYPE II).	
FT	REPEAT	359	430	C (TYPE II).	
FT	REPEAT	431	505	D (TYPE I).	
FT	REPEAT	506	577	E (TYPE II).	
FT	REPEAT	578	652	F (TYPE I).	
FT	REPEAT	653	724	G (TYPE II).	
FT	REPEAT	725	799	H (TYPE I).	
FT	REPEAT	800	874	I (TYPE I).	
FT	REPEAT	875	949	J (TYPE I).	
FT	REPEAT	950	1021	K (TYPE II).	
FT	REPEAT	1022	1093	L (TYPE II).	
FT	REPEAT	1094	1165	M (TYPE II).	
FT	REPEAT	1166	1180	TYPE I (INCOMPLETE).	
SEQUENCE	2249 AA; 224333 MW; A9D6646C089DF087 CRC64;				

Query Match		5.5%	Score 642.5;	DB 1;	Length 2249;
Best Local Similarity		23.6%	Pred. No. 7.0e-16;		
Matches		522;	Conservative 210;	Mismatches 776;	Indels 707; Gaps 114;
Qy	261	ALGADTRVDDYGYALGYSOILNNNNNNNNKAVYPEGNGSNIKSKATGNGLFESISST	320		
Db	36	ALGVATG-----IATNNAAFSNVGNWNNNEITAAGVANGTPAGGPQNNAFTYGGD-	89		
Qy	321	IKRKIIINVAGYEDTDVNA-----OLKAIVENLAKRQITPKGDDNGTVGKKLGEPLTI	375		
Db	90	---YTVTADAADRIIKAINVAGTTPVGLNITQNTVGSIIKGG--NLLPVLNAGKSLTL	144		
Qy	376	KGETQADKLPDNNNIGVYVDTNNTGL-----KVKLAKNL--SCLE	413		
Db	145	GNNAVA-----ANHFADAPADNYTGLGNIALGGANAALIIOQAAPSKITLAGNIDGGII	200		
Qy	414	TVSTRK-----NLTASEKVTYVGSNTAELOSGG-----LTFTPTTNASTDKTVYGTG	461		
Db	201	TVKTDAAALNGTIGNTNALATVNVGAG--TATL--GGAVIKATTTKLTNAASVLTL--	251		
Qy	462	LKFTDNSNT-----ALEDTT-----RITKDKIGFSNKAGTVDENKPYLKD	502		
Db	252	---TNANAVLTGAIDNTGTDGNVGLNGLNALSQVTD-IGNTNSLATI-----	296		
Qy	503	KLKYGNSLTNNGLTVNNTIGGSKQIQVGADGIKFADVNV-----NVSNAK--	551		
Db	297	--SVGAGTATLGGAVIKAT--TTKLTDAASAVKFTNPVVVTGAIDNTGNANGIVTFT	350		
Qy	552	GTTRITEEIGFADAGKVKKSPYLDKKQLQVGGVKITKDSGINAGDOKISNVKDATDD	611		
Db	351	CNSTVT--GNVNTNALATVNVGAGL---QVGGVVK-----ANTINLTDN	392		
Qy	612	TDVITYKOLKVOODADGALQSFSIRDEKQGEFTISNLSNGNTPTFTITFAGEN---	668		
Db	393	ASAVTTPVVV---TGAID-----NTGNANGI--VTFETGNSVT	428		
Qy	669	-GISISNDIAKGVKVGIDPTNG-----LTPKPLVGSDDKOKTQLVIEQVASGNDTKNII	723		
Db	429	GDIGNTNALATVNVGAGTATLGGAVIKATTTKLTNA--	465		
Qy	724	RGLSPTLPSITNAGVTRTEOGNTITSDEKSKAASIGDILNTGFNLKNNNSVGEVSTY	783		
Db	466	-----SVLTLTNANAVLTGAIDNTTGGD-----NVG-VLNLNALSQVGTGNTNSL	512		
Qy	784	NTVDFIDG-----NATTAKVT-----YDETNOTSK--VTVYDNNV	815		
Db	513	ATISVAGTATLGGAVIKATTTKLTDAASAVKFTNPVVVTGAIDNTGNANGIVTFTGN-	571		
Qy	816	DEKTIELTGDKGKTKIGVKTTLTNTNANGKATNFSTDDNALVNKADIAENLNTLAKE	875		
Db	572	---STVTGDIIGNTNSLATISVAGTATLGGAVIKATTT---KLTNAASVLTLTNANA--	622		
Qy	876	IHTTKGTADTALOTFKVKKDGATDDETTIVGKDGQTONGKTVNTLKLGE-----NGLTVAT	931		
Db	623	-----VLTGAIDNTGTDGNVGLNGLNALSQVTDIGNT	656		
Qy	932	NKDGTVTFEINTQSGLKAGDSTTLNKGSLSKNPNASNEQIQVGADGVKFAK-----VD	984		
Db	657	NSLATISVG-----AGTATL--GGAVIK--ATTTKITNNAVSAVRFTNPVVVTGAID	703		
Qy	985	K-GNSSTGI---DGFSTRITKQIQGTGANGSIDTTKPLTKDKLKGVEVLTNTGINAGG	1040		
Db	704	STGNANGIVTFTGNSVTGID-IGNTNA-----LATVNVGAGTATLGGAVIKATT	752		
Qy	1041	KKITNIQSGDITQNSN-----DAVTGG---RVYDLKLETSKINSAAKTAQNSLHEFS	1090		
Db	753	TKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGLNGLNALSQVTDIGNT--NSLATIS	810		
Qy	1091	VADEOGNIFTVSNPSSYDTSK---TSDVITTFAGENGITTKVNGVVRVVGIDQTKGLTTP	1147		
Db	811	VG---AGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTG-----AVDNTTGT---	855		
Qy	1148	KLTVGNNGNKGIVTDSKDCQNTIT---GLSNTLANVTNDGAGHALSQGL---ANDTKTR	1201		

Db	856	---GDNVG---VLNLNALSQVTDIGNTNSLATI--SVGAGTATLGGAVIKATTTKLTN	907		
Qy	1202	AASIGDVLNAGFNLOGNEAVDFVSTYDVTDFIDGNATTAKVTYDDTSKTSKVYDVNVND	1261		
Db	908	AAVLTTLTNANAVLTG---AIDNTTGGDNGVGLNGLNALSQVTDIGNTNSLATISVGAG	964		
Qy	1262	NKTEVTSRDKLG---VKTTTLTKTSANGNATKFS-----AADGALVKASDIA--THL	1310		
Db	965	TAT-----LGGAVIKATT--TKLTDAASAVKFTNPVVVTGAIDNTGNANGIVTFTGN	1015		
Qy	1311	NTLAGDIQTAKGASOASSASYVDADGNKVYDSTDKKYQVNDKGGQVDKKNKEVAKOKLV	1370		
Db	1016	STVTCNVGNTNALATVNVGAGLLQVQ-----GVVKANTINLTNAS	1057		
Qy	1371	AOAQTDPGTLAQMNVKSVINKEOVNDANKKOGINEDNAFIKGLENAAKDTTKNAAVT--	1428		
Db	1058	AVTFT-----NPPVVVTGAIDNTG--NANGIV-----TFTGNSVTGNN	1093		
Qy	1429	VGDLNVAQTPLTJFAGDTGTAKKLGETLTIKGG--QTDTNKLTDN-----	1472		
Db	1094	VGNTNALA-----TVNVGAGLLQVGGVVKANTINLTNASAVTFTNPVVVTG	1141		
Qy	1473	-----NIGVVAGTDTGVKLAOKDLTNLN---SVNAGGTTRIDEKGLSFFVDANGQAKAN	1521		
Db	1142	AIDNTGNANGIVTFTGNSVT--TGDIGNTNALATVNVGA-----GIT--LOAGGSUAAN	1192		
Qy	1522	T-----PVLISANG--LDLGGKRISNI--GAAVDDNDA---VNEKQFNEVAKTVNNLNQOS	1569		
Db	1193	NIDFGARSTLEFNGPLDGGKAIPIYPFKGATANGNALLNVTNLTTLTASHLTIGTV-AEI	1251		
Qy	1570	NSGASLPFV-----VTDANKPINTDGPQKKAIGADGKYYHANANGVVPVOKDGKPI	1622		
Db	1252	NIGAGNLFTIDASVGDVTILNAQIN-----FRARDSVLVLSNLTGTVGN---NIL	1299		
Qy	1623	TDADKLANLAHG-----KPLDAGHOVVASL-----CGN-----SDAI	1655		
Db	1300	LAADLVAPGADGTVFVNGVNGVNGVNSVAGTARNITGCGGNKFNFTLLIYNAVITITDDV	1359		
Qy	1656	TLTNIKSTLPIDTPTNGNAGQAQSLPSLSAAQSSNAASVKDVLNVGNFLQTNHNOVD	1715		
Db	1360	NLEGIONVLI-----NKNADFTSTAFNAGAIQINDATYTTIDANN--NLNIPAGNIQ	1410		
Qy	1716	PVKAYDTVNFVNGTICADITSVRSADGTHMSNTVNTALAAATDDGNNVLKAKDGKPYKADD	1775		
Db	1411	FAHADAQLVLQNSSGNDRTITLGANIDPDN-----DDEGIVIL-----	1448		
Qy	1776	LMPNGSLKAGKSASDA--KTPTGLSLVNPKNAGKSGTDAVALNNLSKAVFKSKDQTTTT	1833		
Db	1449	---NSVTAGKKLTAGGKTEGGAHKLOTILFKGA-GDC-----STAGTTFT	1491		
Qy	1834	VSDGISIQGKDNSSITLSKDG--LNVGKGVISNVYKGTKDTDAANVQQLNEVRNLLGLGN	1892		
Db	1492	-----TNIVLDITGQLELG-----ATTANVLFNDVAVOLTOGN	1525		
Qy	1893	AG---NDNAD-----GNQVNIADIKDPNSSSSNRNRIKAGTVLGCKGNNDTEKLA--	1941		
Db	1526	IGGFDFNFKANGMYTLNNNVNAGAVQ--NTGGTNGNGTFLI-----VLGASNLNRVNGIAML	1579		
Qy	1942	-TGGVQGVQVDKDGNNAN--GDLSNVVWVKTKQDGSKRALLATYNAAGQNTYLTNNPAEADRI	1999		
Db	1580	KVGAGNVTIAKGGKVKIGEIOGTGTNT-----LTLPAHFNLTG-----SINKT	1622		
Qy	2000	NEQIRFHVNDGNOEPVYVQGRNGIDSSASGKHSVAIGFOAKADGEAAVAIGROTQAQNG	2059		
Db	1623	GGQALKLNFMMNGGSSGVW-----GTAAANSVGDITTAGAT	1657		
Qy	2060	SIATGDNAAQTGDOSIATIGTGVNVTGKHSGAIGDPTVTKADNSYSVGNNGNFIDATQTDV	2119		
Db	1658	SPASVSNAGKATLGGTTSFANTFT--NTGAV-----TLAKGITSFAKN---VTAISF--	1706		
Qy	2120	FGVGNNTIVTESNVAJGNSAISAGTHAGTQAKKSDGTAGTTTTAGTA-----TCTVKGF	2174		

Db 1707 --VANSATINFSNLAFSN--ITGG-----GTTLTGANOVTYGT-GSF 1747  
QY 2175 AGTAVGAVSGA--SGABRRIONVAAGEVSATSTDAVNGSOLYKATQGIANATN 2227  
Db 1748 TDTLTNTTFDGAAGSGGNILIKSGSTLDSGVSTLAL-----VVTATN 1791

RESULT 2  
YDBA\_ECOLI  
ID YDBA\_ECOLI STANDARD; PRT; 2003 AA.  
AC P3366B; P76087; P76088; P76856; P76857; P76859;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL 205.9 KDA PROTEIN IN PAAY-ACPD INTERGENIC REGION.  
GN YDBA OR B1401/B1405.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Itoh T.,  
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,  
RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,  
RA Takeuchi K., Takeuchi Y., Wada C., Yamamoto Y., Horiiuchi T.;  
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 28.0-40.1 min region on the linkage map.";  
RL DNA Res. 3:363-377(1996).  
RN [3]  
RP SEQUENCE OF 464-2003 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=92190338; PubMed=1665988;  
RA Moszer I., Glaser P., Danchin A.;  
RT "Multiple IS insertion sequences near the replication terminus in  
Escherichia coli K-12.";  
RL Biochimie 73:1361-1374(1991).  
CC -!- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYSG (AC P25928).  
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR  
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT  
BETWEEN AMINO ACIDS 839 AND 840.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: AE000237; AAC74483.1; ALT\_SEQ.  
CC EMBL: AE000237; AAC74487.1; ALT\_SEQ.  
CC EMBL: D90778; BAA15009.1; ALT\_SEQ.  
CC EMBL: D90778; BAA18880.1; ALT\_SEQ.  
CC EMBL: D90779; BAA18881.1; ALT\_SEQ.  
CC EMBL: X62680; -; NOT\_ANNOTATED\_CDS.  
CC Ecogene: EG11307; ydbA.  
KW Hypothetical protein.  
FT CONFLICT 489 489 I -> V (IN REF. 2).

FT CONFLICT 495 495 I -> V (IN REF. 2).  
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;  
  
Query Match 4.8%; Score 556.5; DB 1; Length 2003;  
Best Local Similarity 22.4%; Pred. No. 7.5e-13;  
Matches 492; Conservative 243; Mismatches 776; Indels 685; Gaps 120;  
  
QY 361 NGTGVKKKLGELTIKGETQADKLTDDNNNIGVVTNNNTTKYKVL-----AKNLSGLE 413  
Db 143 NNDVILDKTEKTLTIR---DSVFYTENADGTISLSDSNRGRKATINLWQIDEANNTVALE 199  
QY 414 TVSTKNLTASEKVTGSGNNTAELOSGGLTFTPTTNASTDKTIVYGD---GLKFTDINSNTA 471  
Db 200 GVS-----ADGATKWQYHNGELVITGDNATYNNNGKT--TVDGKSDTSTEINGNGKV 251  
QY 472 LED-----TTRITKDKIGFSNKAGTVDENKP-----YLDKDKLKV---GNSLNN 513  
Db 252 IQGDLVDVSGGGHGIDITGDSATVDNK-GTWTVTDPESMGIQIDGDKAIYNNNGESTITN 310  
QY 514 GG-----LTVNNT-----IGGSNKQ-----IQVGADGIGKFAVDN 542  
Db 311 GGTGTQINGDDATANNGKTTVDGKDSGTGTGTEINGNGKVIQDGDLDVSGGGHGIDITGDS 370  
QY 543 VNVSNAAKFTTTRITE-BEIGFADADGKVDKSPYLDKKQLQVGG---VKITKDSGI-N 596  
Db 371 ATVDNK---GTMVTDPEISG-----IQVDGQAVVNNNEGESAITN 408  
QY 597 AGDKISNVKDATDDTDAVTYKOLKQVQDADGALQSFSIRDEKGOEFTISNLSYNGNTP 656  
Db 409 GGTGTQINGDDATANNGKT-----TVDG-----KDSGTGEIA---GNNGKVI 448  
QY 657 NTFETITFAGENGISISNDIA---RGKVKVGIIDPINGLTPTKLTGVSDDKGTQLVIEQ 712  
Db 449 QDGDLDVSGGGHGIDITGDSATVDNKGTWV-----TDPE-SIGIQIDG-----DQ 493  
QY 713 VASGNTKNIIRLSPLTSITNAGVVTPEQGNNTTSDDEKSKAASIGDILNTGPNLKN 772  
Db 494 AIYNNNEGS-----TITN-GGTGTQINGDDATANNGKTTVDGKD--STGTKIAG 540  
QY 773 NSNSVCFVSTYNTVDFIDGNATTAKVYDETNQTSKVYDVNVDEKTIETLTGNGKTN-- 830  
Db 541 N---IGIVN-----LDGSLT-----VTGGANGVENIGDNGTVNNK 572  
QY 831 ---KIGV---KTTTLTTT---NANGKATNFSTDDNALVN-----A 862  
Db 573 GDIVVSDTSGIVGLINGEGATVNTGDVNVSNENATGESITNSGKVSLSAGSMQVGFSTG 632  
QY 863 KDTAENLNTL---AKEIH-----TTKGTADTAIQT--FKVKKDGATD-----DE 901  
Db 633 VDLNGNNNSVTLAAKDKLVVGGQKATGINVSGDANTVNTGNVLVDKDKTADNAAEYFFDP 692  
QY 902 TITVKGDTGQNGTAVNTLKLKGENGLTVATNKDGTVTFGINTOSGL-----KA 949  
Db 693 SVGINVYGSNNVNTL---DGKLTVSDSEVT-----SRQSNLFDGSAEKTSGLVVI 740  
QY 950 GDSTTLNKDG---LSIKNP-ASNEIQVGADGVKFAK---VDKGNSTGIDGTSRITKD- 1001  
Db 741 DGGNTVMNMGLELIGKKNALADGSOVTSRTGYSYTSVIVVSGESSVYLNGLDITISGEF 800  
QY 1002 QIGFTG-----ANGS-----LDTTPHPLTKDKLVGEVEITWTGIN-AGKKK 1042  
Db 801 PLGFAGVIRVQDKALLEIGSGATLTMDIDISFEHHGTR-----TVEIQNLGFAFVTGEN 854  
QY 1043 ITNIQSGDIT--ONSND-----AVTGRVYDLKTELESKINSAAKTAQSL-HEFS 1090  
Db 855 TTGINSGTISLQNGKDPAPSPVILLATNGGSSATNAGT-ITGKV-----TEQHSVFPNKS 908  
QY 1091 VADEOCHHTFTVSNPYSSYD-----TSKTSDDVTTFAGENG I-----TTKVNKGV 1133  
Db 909 TG--TSNSFFIENDVSSITGLVAQSNSTIINTDSGIIDLYGRGSGVGLMADSTAEANOQK 966  
QY 1134 VRVG---IDQTKGLTTPFKLTVGNNN-----GKGIVIDSKDQNTITGLSNTLANVT--- 1181



Db 967 ITLDSMMWDAND--TTAMRDIASNSAIDFGTGVGVGTDYSYSGAGKNATAINQLGGVITIY 1024  
Qy 1182 NDGAGHALSOGLANDTKTRAASIGDVLNAG-ENLQNGEAVDFVSTVDTDFIDGNATT 1240  
Db 1025 NAGAGMA-AYGASN-----TVINQGTINLERNG-----NYD--DSLAANTLV 1063  
Qy 1241 AKVTYDDTSKTSKVYVDVNDKNTIEVTSDDKGLGVKTTTLTKTSAN-GNATPKFSAADGDA 1299  
Db 1064 GMAYVEHGTAINDOTGVINIVGTGQAFYNDGTG-----TIVNYGTICTFGWCQSGN 1115  
Qy 1300 LVKASDIATHLNTLAGDIOTAKGASQASSASYVDADGNKVYIDSTDKKYYQVNDKGOVD 1359  
Db 1116 EYNTDDFTSLYITGGDPTITRSGTETVLNLSAAV-----TDKLAGNVVNSGTLS 1164  
Qy 1360 KNEVAKDKVAQAQTPDG-----TLAQMNVKSVINKEQVNDANKKOGNEDNAFIKG--LE 1414  
Db 1165 GDQITVSSGLL--ENTSGGIINLVKLDKGAVI-----KNAGVMTNNVDVSGGILN 1213  
Qy 1415 NAAKDTK--TKNAAV-----TVGDLNVAQPTLTFAGDTGTTAKKLGELTILIKGGQDPT 1466  
Db 1214 NAGEMTAQITNAGADSLVNTGTINKIVQAGVFNNSGVTGRMMS-----AGGV--F 1266  
Qy 1467 NKLTDNNIGVAGTGGFTVKLAKDLTNLNSVNAAGTRIDEKGISFVDANGQAKANTPVL 1526  
Db 1267 NNOITGALMRGAALTGTAVANNEGTVNLGSSSEGN-----NTGMLE 1307  
Qy 1527 ANGLDLGGKRISGAAVDDND--AVNEFKQFNEVAKTVN-MLNQSNQSGASLPFVWVDAN 1583  
Db 1308 VN-----NNSAFNRRGEFLDNDKNAVHINQSGTLYNTGHMINSSSHNGA---VNMWGGN 1360  
Qy 1584 GKPIN-GTIDGKPPQAI-----KGADCKYYHANANGV-PVDKDGKP-----ITDAD 1626  
Db 1361 GRFINDETIDVSAKSLVYSANNAGDQNAFFWNQDGNFINFDHDSASAVKVFTHSFIAQND 1420  
Qy 1627 KLANLAHKGKPLDAGHGVVVASIGNSDAITLTNIKSTLPQIDTPTGNANAGQASLP 1686  
Db 1421 GIMNISGTG-----AVAMEGDKNA-----QLVNNGTINLG----- 1450  
Qy 1687 SAAQSNAAKSKVDLVNGVFNLTQHINOVDFVKAYDTVN-FVNGTGADITSVRSADGTMSN 1745  
Db 1451 -----TAGTTDTGCMIGMQLDANATADAVIENNGTINIFANDSFA-----FSVLGTGVH 1498  
Qy 1746 ITVN-TALAATDDGDNVLKAKD-----GKFYKADDLMPN----- 1779  
Db 1499 VVNGTYYIADGVGTSGGLIKQGDGINVEGMNNGNSSEVHYGDTLTPDVKPKNTSVTS 1558  
Qy 1780 GSKAGKSADAKPTGLSLVNPAGKSGTGDA---VALNLSKAVFKSKDGTITTTVS 1835  
Db 1559 GSDAGGSMNLNGYVVGTVNNGSAGKLKVNNASMNGVEINTGFTA-----GTADTTVS 1612  
Qy 1836 SDGISIOGKD-----NSSITLSKDLNVLGVGKVISNVGKGTKDTDAANVOQLNEVRN 1886  
Db 1613 FDNV-VEGSNLTDAADATSTSVVWTAKGSTDASGNVDVMTSMKNAY-TDVATDASVNDIAK 1670  
Qy 1887 LLGLGNAGND-----NADGNQVNIADIKKDPNSGSSSNRTVYKAGTVLGGK 1933  
Db 1671 ALDAGYTNNELFTSLNVGTTAEALSALKQV-----SGSQAT-TVFREARVL----- 1715  
Qy 1934 NNDTEKLTATGCVQGVQVDKGDGNANGDLSNVVVKTKQDKSKALLATYNAAGOTNYLTNPA 1993  
Db 1716 SNRFSMLADAAPKVG-----NGLAFNVAK-----GDPAELG-----NN-- 1750  
Qy 1994 EADRINEQGRFFHVNDGNQEPVVOGRNGIDSSASGKHSVAIGFQAKADGAAVAIGRQ 2053  
Db 1751 -----TEYDMLALRKITIDLESQTMISLEYGI-ARLD----- 1780  
Qy 2054 TQAGNOSTAIGDNAQATQDQSIAGTGNVYTGKHSIGAIGDPTVKADNSYSVGNNNQFID 2113  
Db 1781 -----GDGAQKAGD-----NGVTGYGSQFGLKHQMSFGMMNWN----- 1816  
Qy 2114 ATQTDVFGVGNNTVTRBSNSVALGNSAISAGTHAGTQ-----AKKSDGTAGTTT- 2164

Db 1817 ALRYDVNNLDS-----SRSTAFG-NTNKTADTVDKQOYLEFRSEGAKTFFPSBGLKVTP 1869.  
Qy 2165 -AGA--TGTVKGFAGQAVGAVSVGA-SGAERRIQNVA-----AGE-----VSATSTDAV 2210  
Db 1870 YAGVKLRHTLEGGYQERNAGDFNLNMNSGSETAVDSIVGLKLDYAGKDGWSASATLEGGP 1929  
Qy 2211 NGSOLYKATQCIANATNELDHRHONENKANAGISS 2246  
Db 1930 NLS--YAKSQRTASLAGAGSQHFNVDDQKGGGINS 1963  
RESULT 3  
OMPB\_RICJA STANDARD; PRT; 1656 AA.  
AC O06653;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).  
GN OMPB.  
OS Rickettsia japonica.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=35790;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-YH;  
RA Uchiyama T.;  
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia  
RT japonica."  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
CC SIMILARITY).  
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
CC (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
CC LAYER WITH HEXAGONAL SYMMETRY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB003681; BAA20138.1; -  
KW Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.  
FT CHAIN 1 1338 32 KDA BETA PEPTIDE.  
FT CHAIN 1339 1656 POLY-GUY.  
FT DOMAIN 528 533  
SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;  
-----

Query Match 4.2%; Score 492.5; DB 1; Length 1656;  
Best Local Similarity 22.7%; Pred No 1.1e-10;  
Matches 419; Conservative 198; Mismatches 690; Indels 539; Gaps 93;

Qy 656 PNTFTTITFAGENGISINDIAKGVKVGIDPINGLTTPKLTGSKDKGKTLVIEQVAS 715  
Db 68 PNA--VITANANNGINLTNTPAGS-----FNGFLF-----SNANLAVTVSEDTTL 110  
Qy 716 G--NDRKNIRGLSPTLPSITNACGVRT-TEQGNITVTSDEKSKAAASIGDILNTGFNLN 772  
Db 111 GFINNAANNANRFLTL----DAGKTLTITGGITNVQSAATHNAQNTIVAKFNGGAIA 166  
Qy 773 NSNSVGFVSTVNTVDFIDGNATTAKVTYDETNTQTSKVYDVNVDEKTELTDGDKTNK- 831

Db 167 ND-----LSGLGIDTF---GAAASTLVFLDLPNTTQKAPLILADNALI-VNGANGTLNVT 217  
QY 832 ---IGVKTTLTTTNA-----NGKATFSTDDNDALYNKADIAENLNTLAKEIHTTKGTAD 884  
Db 218 NGFIQVDSKFATVKAINGIDGGGEMFNATNA--NALNLQAGGTTI--NFNGTDTGTR 273  
QY 885 TALOTFKVKKGATDDETITVKGDTQNG-KIVNTLKLKGENGLTVATWKNKGTVPFGINT 943  
Db 274 LVL-----LSKNGAATDFNTVSGGLKGIIEIENLTVAI---NGQLIANAGPANAVIGTNN 326  
QY 944 QSLKAGDSTTLNKGSLIKNPASNEQIQVGADGVKFAKVDKGNSSSTGIDGTSRIKDOI 1003  
Db 327 GAGRAAG-----FVSVDNKGKAAT-IDGQVYAKDMVI 357  
QY 1004 GFTGANGSLDTTKPHLTKDKLKVGEIETNTGINAGGKKITNIQSGDITQNSDAVTGGR 1063  
Db 358 QSANANG-----QVNERHIVDVGIDGTAFKTAASIVAITQNSNEFTT--- 400  
QY 1064 VYDLKTELESKINSAAKTAQNSLHEFSVADEQ---GNHFT---VSNPYSSYDTSKTSIDVIT 1118  
Db 401 -----DFGNLAAQVTPDPTMTLTGN-FTGDANPN-----GNTAGVIT 436  
QY 1119 FAGE-----NGITTKVNVKGVVRVGDIDTKGLTTPKLTVCNNN-----GRK 1158  
Db 437 FRANGTLASADANVANVTNNITAIIPASG---VGVVQLSGTHTAEURLGNAGSVFKLAG 493  
QY 1159 IVIDSK-----DQONTITGLSNTLANVTNDGAGHAL-SOGLANDTDK 1199  
Db 494 TVINGRVNQTVLVGGVLAAGAITLDGSATITG-----DIGNGGGGAALQSITLANDATK 547  
QY 1200 TRAASGCVNLNAG---FNLQNGEAVDFVSTVD--TVDFIDGNATTAQVYDDTSKTSKV 1254  
Db 548 TLTGLGANIISANGGTFINFGANGGTTIKLTSTQNNIYVDCDLAIAITDQTVGVVDASSLT--- 604  
QY 1255 VYDVNVNDKTIETVSDKLGKVTTL-----TKTSANGNATKESAADGDALVKASDIA 1307  
Db 605 -----NAQTLTISCTIG-IIGANNFTILGQNFICSSKTTLAG-----GNVAINELVIG 650  
QY 1308 THLNTLAGDIQTAKGASQASSASVYDADGNKVIYDSTDKKYVQVNDKQGVKNKEVARD 1367  
Db 651 NN-----GSVQFAHNTYLITRTTN--AAGGKIIFNPV-----VNNNTTLAAG 691  
QY 1368 KLVAQAOTPDGTLAQMNVSVINKEOVNDA--NKKOGINEDNAFKGLENAAKDFTKNA 1425  
Db 692 TNLGSAANP---LAEINFGS---KGARADTVLNVGEGVN-----LYATNITTTDANVGSE 740  
QY 1426 AVTVGDLNAVAOTP-----LTFAGDTGTTAKKLG-----ETLTIKGGOT 1464  
Db 741 VFNAGKNIVSGTVGCGQGNKFNVALDNGTTVKFLGNATFNQNTTIAANSTLIQISGNT 800  
QY 1465 DTNKLTDNNIGVV--AGTDGFTVKLAKDLTNLSV-----NAGGPRIDEKGISFVDANG 1516  
Db 801 ADPIASADGTGIVEFVNTGFINVTLNKQAVPVNALQKITVSGPGNVVVNEIG---NAGN 856  
QY 1517 QAKANTPVLISANGLDLGGKRISNIGAAVDNDADVNFQFNEVAKTV--NNLNQNSCASL 1575  
Db 857 YIGAMTDITAFENSSLGA---VLFLPSGIPFNDA-----GNTIPLTIKTSVGNETAEGFSV 909  
QY 1576 PFVVTDANGKPGTGDKPKQKAIKAGDGKYYHANANGVPVDKGPITDADKLANLA--- 1632  
Db 910 PSVI-----VSGVDS-----VIADGQVIGQONNIVGLGLGS 940  
QY 1633 -----AHGKLPLDAGHOVVASLGGNSDAITLNTIKSTLPQIDTPNTGNAGANAQSLPSSL 1687  
Db 941 DNGIIVNATTLVAG---IGTINNOCGTVILSG-----GVNTPGTIVYGLG---TGIG 986  
QY 1688 AAQOSNAASVKDVLNVGNFNLQNHNOVDVKA-----YD---TVNFEVNGTG 1730  
Db 987 ASKFQVOTFTDYNLNLGNIATNTINDGVTVTTGGIAAGGIAGTDFDCKITLGSVNGN- 1045  
QY 1731 ADITTSVRSADGTMSTNTVTALAAATDDGCVNLIKAKDGKF-YKADDLMPNGSLKAGKSAS 1789  
Db 1046 ---ANVRFADGIFSN---STSMIVT-----TKNNGTVTVYLGNAFVGN-----IG 1084

## RESULT 4

HLVA\_PROMI  
ID HLVA\_PROMI STANDARD; PRT; 1577 AA.  
AC P16466;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-NOV-1990 (Rel. 16, Last annotation update)  
DE HEMOLYSIN PRECURSOR.  
GN HPMA.  
OS Proteus mirabilis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Proteus.  
OX NCBI\_TaxId=584;  
[11]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.  
RC STRAIN-ISOLATE 477-12;  
RX MEDLINE=90170827; PubMed=2407716;  
RA Uphoff T.S., Welch R.A.;  
RT "Nucleotide sequencing of the Proteus mirabilis calcium-independent hemolysin genes (hpmA and hpmB) reveals sequence similarity with the Serratia marcescens hemolysin genes (shla and shlb).";  
RL J. Bacteriol. 172:1206-1216(1990).  
CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY DEFINED.  
CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HPMA REQUIRES HPMB FUNCTION.  
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
CC -!- MISCELLANEOUS: THE CONSERVED AMPHIPATHIC DOMAINS IN SHLA AND HPMA MAY BE RESPONSIBLE FOR PORE FORMATION.

CC -1- SIMILARITY: TO S.MARCESCENS HEMOLYSIN (SHLA).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M30186; AAA25657.1; -;  
DR PIR; A35140;  
KW Hemolysis; Toxin; Outer membrane; Signal.  
FT SIGNAL 30 29  
FT CHAIN 30 1577 HEMOLYSIN.  
SQ SEQUENCE 1577 AA; 165869 MW; 175975E0C924B2D9 CRC64;  
  
Query Match 3.9%; Score 450.5; DB 1; Length 1577;  
Best Local Similarity 21.1%; Pred. No. 3.3e-09;  
Matches 389; Conservative 249; Mismatches 605; Indels 603; Gaps 95;  
  
Qy 593 SCINAGQKISNVRDADTDDTAIVYKQKQVQDADGALQSFIRDEKQGEFTISNLYSN 652  
Db 44 SAVNGGTQVINIV---TPNNEGISHNQYQDFNVGKPGAV--FNNALEAGQSQLAGHLNAN 98  
Qy 653 GNTPTNFTETIFAGENGISINDIAKGVKVGIDIPINGLTTPKLTGVGSDKDKCKQLVIEQ 712  
Db 99 SN-----LNQOA-----ASLIINE 112  
Qy 713 VASGNDT-----KNIRGLSPPLPITNAGVTRTEQGNITTSDEKSKAASIGDIL----- 764  
Db 113 VVSRNPSPLLQQQEVFGIAAEY-VLSNPNGITCDGCGFINFS-----RSSLVVGNPLFENG 167  
Qy 765 -NTGFNLKNSN--SVGPVSYNVDFEID-----GNATTAKVYDETNQTSKYTYD 812  
Db 168 QLKGYSTLNNLNLISLG-KNGLNTTGLDLDLAPRIDSRGKITAAEISAFPTQNTFSQHF 226  
Qy 813 VNVDEKTIELGDGKTNKIGVKTTLTTTNNANG-KATNFSTTDNDALVNAKDIAENLNT 871  
Db 227 ILSQKPVSA LDSYFFGSMQSGRIIRIINTAGSGVKLAGKFTADNDLSVK-----ADNIQT 282  
Qy 872 LAKEHTTKGTADTALQTFKVKDGATDDEP-----ITVGKDGTONGTVNTLKLKGENGL 927  
Db 283 -----DSQVRYSDYDKGSENYQNYRGGITVNNSSG--SSQPLTKTELKGNK-- 326  
Qy 928 TVATNKDGTVTFGINTQSLKA---GDSITLTKDGLSIRKNPASNEQIQVGADGVKFAKV 983  
Db 327 -----ITLVASSHNQIKASDLMGDDITLOGADLTI-----DGKQLQ----- 362  
Qy 984 DKNSSTGIDG-----TSRIT--KQOIGFYCA-----NGSLDTTKPHLTCKLKVGEV 1029  
Db 363 ---QKQETIDNRWFYSWKYDYVTEKEQIQOIGSQIDAKNNATLTATKGDVTLDAK----- 415  
Qy 1030 EITNTGINAGKKITNIQSGDITONSNDVATGGRYVDLTKLESKINSAAKTAQNSLHEF 1089  
Db 416 -----INAGNNLAIN-----ANKOIHNGLVKESRSENGKRNHTSRLESGW--- 459  
Qy 1090 SVADQGGHFTVSNPYSDYTSKTSQSDVITTFAGENGITTKVKNKVVYRVGIDQTKGLTT--P 1147  
Db 460 -----SNHQET-ETLKASELT--AGKO-----LGLDAQGSITAQGA 492  
Qy 1148 KLTGVNNNGKIVIDSKQONTITGLSLNTLANVNDGAGHALSOGLANDTDK----- 1199  
Db 493 KLIHANE-----VLNNAKINILNVQKTKNDKTVTDN-----HYMWGGIGGQKNNNNNQOV 545  
Qy 1200 ---TRAASIGDVLNAGFNLOGNEAVDFRVSYDVTDFIDGNATTAKVYDDTSKTSKVYV 1256  
Db 546 SHATQLTADGQQLLAADN-----NVTGSOVKGNQCAFVKTTQ 585  
Qy 1257 DVNVDNKITEVTSKGLGVKTTT-----LPKTSANGNATKFSAADGDALVKASDIATHNL 1313  
Db 586 DVVIDNALSETIS--KIDERTGTAFNITKSSHKNETN----- 621

Qy 1314 AGDIQTAKGASQASSASYVDADGNKVYDSTDKKYYQVNDKQVNDKNEKAVAKDLVAQA 1373  
Db 622 ---QSTG-SELISDAQLTVVSGN-----DNNVIGSLIK-----SADKLGH- 659  
Qy 1374 QTPDGTIAQNMVSVKINKEQVNDANKQGIN-----EDNAFIKGL-----ENAKADT 1420  
Db 660 ---SLGDINVKSAQVTKIKDEKTSIAITGHAKVEDEKQYSAGFHITHTNTKSTSET 714  
Qy 1421 KTKNAAVTVGDLNAVAOTPLTFAG-DTGTTAKKLGETLTIKGG-----OTDNKLKTNNI 1474  
Db 715 EQANSTISGANVDLOANKDVTTFAGSDLKTTA---GNASITGDNVAFVSTENKKTQDNTD 770  
Qy 1475 GVWAGTDGFTVKLAKDLTLNLSNVNAGGTRID---EKGISFYDANGQAKANPPV-----LS 1526  
Db 771 TTISGGFSYT-----GGVDKVGSKADFYDKQHTQTEVTNKGQSQTEVAGDLTIT 820  
Qy 1527 ANGLDL-----GGHRISNIGAAYVDNDAVNFKQFNEVAKTVNNLNQNS 1571  
Db 821 AN-KDLLHEGASHHIVEGRYQESGNIQHL--AVND-----SETSKT-DSLNVGIDV 867  
Qy 1572 GASLPFVVTDANG--KPINGTDPKPKAIKADGADKYYHANANGVYVDKDG-----KPIT 1623  
Db 868 GYNL-----DYSGYTKPV-----KKAIE-----DGVNTTKPGNNTDLTKKVT 904  
Qy 1624 DADKLANLA--AHGKPLDAGHQVVASLGGNSDAITLNIKSTLPDPTPTNGANAGQAQ 1681  
Db 905 ARDAIANLANLSNLETPNVGVEVIGKGGSQSQSDQAVST-----SINAKGIDIDSNN 959  
Qy 1682 SLPSLSAAQSNAAKSVKDLNVGNENLQTNHNQVDFVKAYDTVNFVNG-----TCAD 1732  
Db 960 KLHD-----QGHYOSTQEGISLTANTHTSEATLD--HQHTTTHETKGGQIGVSTKTKGSD 1013  
Qy 1733 ITSRSADGTMSNTVTNTALAAATDDGNNLIKAKDGKPYKADDLMPNCSLAKGASASDAK 1792  
Db 1014 ITVAIKGEGO---TTDNALMET-----KAGSOFTSNGDISINVGENAHVEGAQFD 1061  
Qy 1793 TPTGLSLVNPAG-----KGSTGDAVALNLSK-----AVFKSK 1826  
Db 1062 AQKGTVINAGGDLTLAQATDTHSESQSNVNGSANLKVGTTPESKDYGGFAGNCTTHSK 1121  
Qy 1827 DGTIT---TTVSSDGI-----SIQKDNSS---ITL-----SKDGLN 1857  
Db 1122 EOTAKVGTITGSGIELNAGHNLTLOGTHLSSEODIALNATNKVDLOSASSEHTEKGN 1181  
Qy 1858 VGGKVISNVGKTKDXTDAANVOQLNEVRNLLG-----LGNAGNDNADGN--- 1901  
Db 1182 LSGVQAGFGKKMTD-DASSVNGLSGAQFAIGKQDEKSVSREGGTINNSGNLTINGNSVH 1240  
Qy 1902 ---QVNIADIKDPNSG---SSSNRTVKA--GTVLG--GKGNNDT-----EKLAT- 1942  
Db 1241 LQGAQVNSKDTQLTSQSGDIEITSAQSTDYKNNMGTDIGFMGKKTNTNTPKEVTEKPA 1300  
Qy 1943 ---GGVQGVGD-----KQGNANGDLSNVVVKTKD-----GSKKALLATYNAAGOTNY 1987  
Db 1301 IHNIGKLLVNVEDQKTSHQNTATLETGTLTINSKDLTLSCANVTADSVTGNVGGSLNI 1360  
Qy 1988 LTNPAEADIRINQGIIR--FFHVNDGNQBPVVOGRNIDSSASG-----KHVSVAIGF 2038  
Db 1361 ASQKES---DRHVTGVNVGYNHTNDPKSSQVKN-----TAKAGSLLLEKTIKDTIDSGI 1412  
Qy 2039 QAKADGEAAVAIGRQTOAGNOSIA---IGNQAOTGDOOSIA-IGTG--NVVTGKHSGAI 2091  
Db 1413 KSSTD-----AISDKYNSLSSTIADKICISDETRAKIDQGGFKGVNGIKNIYTC----- 1461  
Qy 2092 GDPSTVKADNSYSYGNNNQFIDATQTDVFGVGNITVTESVALGSGNSAISAGTHAGTQ 2151  
Db 1462 AEGHTANAD-----IKVTHVDNDVATKTTSLTSLNNDLSLVN----- 1498  
Qy 2152 AKKSDCTAGTTTTACATGTVKGFAGQTAAGVAVSGVSGAEBRRIONV 2197  
Db 1499 -----GSTKLTGA---EIVSQQGVDLGGSSV-----KLENI 1527



Db 1558 VLDKDHDLTVTASQADNRSLK---AGHDFTVQAOQIDNSGTMMAAGHADTLKAP-HLRN 1613  
Qy 1218 NGEAV-----DFVSTYDTVDVDFIDGNATTAKVTYDD 1247  
Db 1614 TGQVAGHHIHLINSKLENTGRVDARDIALDVAADFNTGSLVLAHEHDTATLTLAQGTORD 1673  
Qy 1248 -----TSKTSKVVYDV--NVDNK-TIEVTSKKLGK 1276  
Db 1674 LVVDODHILPVAEGTLRVKAKSLTTEIETGNPGLIAEVQENIDNKQAIIVVGKD----- 1727  
Qy 1277 TTTLTKTANGNATKFSADGDALVKASDIATHLNTLAGDIOTAKGASQASSASYVDAD 1336  
Db 1728 ---LFLSSAHGNNV---ANEANALWA-----AGEL-TVKAQNIITNKRAALIEAG 1769  
Qy 1337 GKNVYDSTDKKYYQVNDKGQVDRKNEVAKD-KLVAQAQTDPDGTIAQMNKSVINKEQVN 1395  
Db 1770 GNARLTAIV---ALLNKLGRIRAGEDMILDAPRIENTAKLSGEVQKGVQDVGGGEH-- 1823  
Qy 1396 DANKQGINEDNAFTKLGLENAAKDTKTNAAVTVGDLNVAQTPLTFAGD--TGTAKK- 1452  
Db 1824 --GRWSGICYVNYMLRA--GNGKAGGTIAAPWYGDLTA-EQSLIEVGKDLVYNAGARKD 1878  
Qy 1453 ---LGTFLITKGGOTDNKLTNN--IGVVAGTGTGTVKLAKDLTNLNSVNAAGTR-- 1503  
Db 1879 EHRHLNEGVIQAGGHGHIIGGDVNRVTVVSAMEYEKTEPLPVSLTALD--NRAGLSPA 1936  
Qy 1504 -----IDKRGISFV-----DANGQAK---ANTPVLNAN 1528  
Db 1937 TNFQSTVELLDYLLDONRYEYIWGLYPTYTSEWSYNTLKNLDLGVQAKPAPTAPMPKAP 1996  
Qy 1529 GLDLCGKRISNITGAANDDAVNFKQFNEVAK-----TVNNLNQ 1568  
Db 1997 ELDERGHTLES-----AEGRKIFGEYKKGLEYEKAKMAVQAVEAYGEATRRVHDQ 2047  
Qy 1569 SNSGASLPFVVTWANGKPLNG-----TDGKQO----- 1595  
Db 2048 LCQRYGKALGGMDAETKEVDGIIQFEADLRTVYAKAQDQATIDAEATKVAQRYKSQIDA 2107  
Qy 1596 -----KAIGADGKYYHANANG-----VPVDKDGKPIITDADK----- 1627  
Db 2108 VRLQAIQPGVTVLAKALSAALGADWRALGHSQLMQWRKDFRAGKRGAEATFYFPEQTVLA 2167  
Qy 1628 -----LANLAH-----GKP--LDAGHOVVASLGGNSDAITLTIKSLP----- 1665  
Db 2168 AGAGLTLNGAINGENANAAQNRGPEGLKIGAHATSVSFGFDALRDVGLKRLDIDDAL 2227  
Qy 1666 -----LIRAKDGKF-YK 1772  
Db 2228 AAVLVNPHITRIGAAQTSADGAAGPALAQARQAPETDGVNDARGLSADALASLASL 2287  
Qy 1678 GOAQL-----PSLSAAQOSNAASV-----KDLVNGFN--- 1706  
Db 2288 DAAQGLEVSGRNQAQVADAGLAGPSAAPAAGVADGVPEPTGVDQVQPVVAVGLEQPV 2347  
Qy 1707 -----LOTNHNQVDFKAYDIFNVFNGTG-----ADITSV 1736  
Db 2348 ATVRVAPNAVALPRLFEITRKIDQSKFYGSRYFEQIGYKPDRAARVAGDNYEDTTLV 2407  
Qy 1737 -----RSADGTMSTNV-NTALAA--TDDGNNV-----LIRAKDGKF-YK 1772  
Db 2408 REQVRALGGYERLUPRGVALVAKLMSAGTVGKALGLKVGVAFTAQQLQAQDRFVWY 2467  
Qy 1773 ADDLPNPSLKGAKSADAKPTTGLSLVNPNAKG-----STGDV-----ALNKL- 1818  
Db 2468 VDTVIDGKVLAPRLYLTEATRGQIT--DQYAGGALLIASGDDVTNTDGHIDVSVNGLI 2525  
Qy 1819 -----SKAVFKSKDGTFTTTTTSVSDGISIOGKD-----NSSITLSKDLGNV 1858  
Db 2526 QGRSVKVDAGKGVVAVDSKAGGAGIEADDEVDVSGRDIGEGKLRGKDVRLKADTVKV 2585  
Qy 1859 G-----GKVISVNGKGTKPTDAAVQO-----QINEVRNLGL 1890  
Db 2586 ATSMRYDDKGRLAAR-GDCALDAQGQLHIEAKRLTAGATLKGKVKLVDVDDVK-LGGV 2643

Qy 1891 GNAGNDNAD-----GNQVNIADIKKDPNSGSSNR--TVTKAGTVLGGKNNNDTEKIAT 1942  
Db 2644 YEAGSSYENKSSSTPLGSLFAILSTSTETNQSAHANHYGTRIEAGTLEGKMNLETE----- 2699  
Qy 1943 GG-----VQGVYDKDGNANGDLSNVVVKTKD-----GSKKAL 1975  
Db 2700 GGSYDAHAHTDLSVARDARFKAADFAHAHEKDKVRQLSLGAKVGAGGYEAGFSLGSESL 2759  
Qy 1976 LA-----TYNAAQGTNYLTNNPAEADIRINEQGIREFHVDNQEPVQVRNGIDSSAS 2029  
Db 2760 EAHAGRGMTACAEVKVGYRASHEQSS-----ETEKSVRNANLNFPGGS--VEAGNVLDIGGA 2814  
Qy 2030 GKHSVAIGFQAKADG-----EAAVAIGRQT-QAGNOSIAIGDNAQA----- 2069  
Db 2815 DINRNYGGAAGNAGTBEALMRRAKKVSTKYVSEQTSQSSGWSVEVASTASARSLIT 2874  
Qy 2070 ---TGDSIA--TGTGNVTVTK-----HSGAICDPSTVKAD---NSY 2103  
Db 2875 AATRLGD-SVAONVEDGHEIRGELMAAQAABAETQVLTADTAVALSAGISADFDSSHRSR 2933  
Qy 2104 SVGNNOF-----IDATQTDV-----FGVNNITVTESNVAL-----GSN 2139  
Db 2934 STSQNTQVLGNLSTIATEGDTLVGAKFGGQDQVSLKAAKSVNLMAAEFTFESYSESHN 2993  
Qy 2140 SAISAGTHAGTQAKSDGCTAGCTTTTAGATGTVKGFAGTAVGAVSVGASGAERRIQNVAA 2199  
Db 2994 PHASADANLGAANAVQGVGLGTAGMGTSHTQITNETGTYAG-TSDAANY-----SIDA 3047  
Qy 2200 GFVSATSDAVANGSOLYKATOGIANATNELDHRITHONENKANAGISSAMAMASMPQAVIP 2259  
Db 3048 GKDLNLSGRVRGKHVLDVEGDINATSKQDER---NTNSSGGGWDASAGVA-----IQ 3098  
Qy 2260 GRSMV-----TGGIATH-----NGQGAVALGSLKLS 2285  
Db 3099 NRTLVPVGSAGFNTBHDNSRLTNDCAAGVVASDGLTGHVKGDLNLTGTATADLS--- 3155  
Qy 2286 DNGQWVFKINGSADTQ 2301  
Db 3156 --GRGNLKVGDGAVNAQ 3169

RESULT 6  
OMP\_B\_RICPR  
ID QMPB\_RICPR STANDARD; PRT; 1643 AA.  
AC Q53020; Q9ZCM0;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
GN OMPB OR SPAP OR SPA OR RP704.  
OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiella; Rickettsia.  
OX NCBI\_TaxID=782;  
[1]  
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RP STRAIN=BRINL;  
RX MEDLINE=91045972; PubMed=2122457;  
RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;  
RT "Characterization of the gene encoding the protective paracrystalline-  
RT surface-layer protein of Rickettsia prowazekii: presence of a  
RT truncated identical homolog in Rickettsia typhi";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=BRINL;  
RA Moron C.G., Yu X.J., Walker D.H.;  
RT "Sequence analysis of ompB of Rickettsia prowazekii";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

[3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MADRID E;  
 RX MEDLINE=99039499; PubMed=9823893;  
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria.";  
 RL Nature 396:133-140(1998).  
 [4]  
 RP PARTIAL SEQUENCE.  
 RC STRAIN=BEINL;  
 RX MEDLINE=92114896; PubMed=1370573;  
 RA Ching W.M., Carl M., Dasch G.A.;  
 RT "Mapping of monoclonal antibody binding sites on CNBr fragments of  
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia  
 RT prowazekii.";  
 RL Mol. Immunol. 29:95-105(1992).  
 [5]  
 RP IDENTIFICATION OF CLEAVAGE SITE.  
 RC MEDLINE=92104668; PubMed=1729180;  
 RX Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;  
 RA "Evidence for proteolytic cleavage of the 120-kilodalton outer  
 RT membrane protein of rickettsiae: identification of an avirulent mutant  
 RT deficient in processing.";  
 RL Infect. Immun. 60:159-165(1992).  
 CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION  
 CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.  
 CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M37647; AAA26390.1; ALT\_INIT.  
 CC EMBL; AF161079; AD42234.1; -;  
 CC EMBL; AJ235273; CAAL5140.1; -;  
 KW Antigen; S-layer; Cell wall.  
 FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.  
 FT VARIANT 257 257 V -> A (IN STRAIN BEINL).  
 FT VARIANT 1010 1010 Y -> D (IN STRAIN BEINL).  
 FT VARIANT 1450 1450 A -> S (IN STRAIN BEINL).  
 FT VARIANT 178 179 AA -> VC (IN REF. 1).  
 FT CONFLICT 191 201 TTQEAFLTGA -> INSRSSSVHLVS (IN REF. 1).  
 FT CONFLICT 212 212 T -> I (IN REF. 1).  
 FT CONFLICT 313 313 Q -> L (IN REF. 1).  
 FT CONFLICT 1104 1104 D -> G (IN REF. 2).  
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).  
 SQ SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;  
 Query Match 3.8%; Score 445.5; DB 1; Length 1643;  
 Best Local Similarity 21.4%; Pred. No. 5.1e-09;  
 Matches 392; Conservative 216; Mismatches 713; Indels 509; Gaps 88;  
 QY 776 SVGFV-STYVNTVDFIDGNATKATVYDETQTSKYTYD-VNVDEKTIETLGDNKTNKI 832  
 DB 13 SAGLVASTATIVAGSGVAMGAMQYKRTTNAATFFDGIQD-----QAAGAN-----I 63  
 QY 833 GVKTTTLTTNANGKATNFSTTDNDA-----LVNAKDIAENLNTLAKEIHTTKGTADTALQ 888  
 DB 64 PVAPNSVITANANPIT-FNTPNGHLSNLFLOTANDLAVTINE-----DTTLGFTINIAQ 117  
 QY 889 TFK-----VKRDGATDDE---TIIVKDGDTQ--NGKTVNTLKLKGENGLTVA 930

Db 118 QAKFFNFVAAGKILNITGGITVQEAASNTINAQNALTQVHGAAINANDLSGLSTTFA 177  
 QY 931 -----TNKDCVTTFGIN-----TOSGLKAGDST-----TLNKD- 958  
 Db 178 AAPSIVLEFNILNITTOEAPLTGLANSKIVNGGNTLNITNGFTQVDSNTTFAKITINIDD 237  
 QY 959 --GL---SINKPASNEIOIQVAGDGVKFAKVDKGNSSSTGIDGTSRI-----TKDQIGFT 1006  
 Db 238 CQGLMFNSTPDAANTLNLQVGGNTINF-----NGIDGTCKLVLSKNGAATEFNVT 288  
 QY 1007 GANGSLDTTKPHLTOKLKVGEVEITNTGI-----NAGGKKITHQS 1048  
 Db 289 GTLGG-----NLK-GIIELTAAVAGKLIISOGGAANAVIGTDNGAGRAAGFIVS 336  
 QY 1049 GDITQNSNDVAVTGRVYDVKLELESKINSAAKTAQNSLH-----EFSVADE-----QG 1096  
 Db 337 VD--NGNAATISQVYAKNMVQSANAGGOVTFEHIHVDVGLGTTNFKTADSKVIITEN 393  
 QY 1097 NHFTVSNPYSSYDT-----SKTSDVITFAGENGI---TTKVNGK 1132  
 Db 394 SNFGSTN-FGNLDTQIVVPDPTKILKGNFIGDVKNNGNTAGVITFNANGALVSASTDPNIA 452  
 QY 1133 VVRV-----GIDQTKGLTTPKLVGNNN-----GKGVIDSKDGQNTITGLSNTLA- 1178  
 Db 453 VTNINAIEAGAGVWELSGTHIAELRLGNGSGIFKLDAGTVINGPVPVQNALMN-NNALAA 511  
 QY 1179 -----NVNDGAGHALSQ-GLANDTDKTRAAS-----IGDLNAGNLOQNGE 1220  
 Db 512 GSTLDGSAIITGIDGNGVYNAALQHTLTLANDASKILALDGANIIGANVGGAHFQANGG 571  
 QY 1221 AVDFVSTYDFVDFIDGNATKATVYDDTSKTSKVYVDVNVNDKTIETVSDKLGKVTTL 1280  
 Db 572 TIKLTNTQNNI-----VVNFDLDTTKGVVDASSL 603  
 QY 1281 TKTSANGNATKFAADGDALVKASDIATHLNTLAGDTQAKGASQSSSASYVDADGNKV 1340  
 Db 604 T-----NNQTTINGSIGTVVANTKTLLA-QLN--IGSKTILNAGDVAINELVNIENGSVQ 656  
 QY 1341 IYDST--DKKYQVNDKQGVQDKNKEVAKDLVAQAQPPDCTLAQMNVKSVINK-EOVND 1397  
 Db 657 LNHTVLTITKINAANGQI---IVNADPLNTNTLADGT-----NLGSAENPLSTIIFA 708  
 QY 1398 NKKOGINEDNAFIKGLENAAKDTKTKNAAVTVGDLNVAQTPLTFFAG----- 1444  
 Db 709 TKAANADSIILNVGKVNLYANNITTND--NVGSLHFRSGGTSIVSGVGGQGHKLNNL 766  
 QY 1445 --DTGTTAKKLGETLTIKGG-----QTDINKLTD-----NNIGVV--AGTDGETVK 1486  
 Db 767 ILDNGTTFVKFLGDT-TFNGGKTIEGKSILQISNNYTTDRHVESADNTGTLEFVNDPTIVT 825  
 QY 1487 LAK-----DLTNL-----NSVN-----AGGTRI 1504  
 Db 826 LNKQGAIFGVKQVIISGPGNIVFNEIGNVGVHGAANSISFENASLGTSLFPGSTPL 885  
 QY 1505 DEKGISFVDANGQA-KANTPVLISANGLDLGGKRISNIGAAVDDND---AYNFKQFNEVAK 1560  
 Db 886 DVLTIKSTVGCTGVDFNAPIVVVSIGID---SWINNGQIIGDKKNIALLSISDSDNITV 941  
 QY 1561 TVNNL-----NNQNSGASLPFVVYTDANGKPIGTGCKPQKAIKGADGKYYHAN-ANGVP 1614  
 Db 942 NANTLYSGIRTTKNNQG-----TVTLSGGMPNN-----PGTIYGLGLENGSP 983  
 QY 1615 VDKDGPITDADLANLAAGHKPLDAGHQVVASLGGNSDAITLT-----NIKSTLP 1665  
 Db 984 KLQVTFDTDDYNNLSGLIANNVTI-----NDIVTLTTGGIAGTDFDAKITLG 1030  
 QY 1666 QIDTPNTGNANAGQAQSL-----PSLSAAQQAASNAASVDLVNGLVFNQTNHINQVDFKAYD 1721  
 Db 1031 SVN---GNAVRFVDSFTSDPRSMIVATQANKGT---VTYLGALVSNIGSLD-----T 1078  
 QY 1722 TVNFVNGTGDADITSVRSADGCTMSNI-----TVNTALANTDDDDGNVLKAKDGKFKYA 1773

```
Db 1079 PVASVRFTEGND-----SGAGLQGIYNSIDFGTYNLTIL-----NSNVILGGGTAINGE 1129
Qy 1774 DDLMPNGSILKA-GKSASDAKPTGLSLVNPAGKSTGDAVALNNLSKAVFKSGDGTITT 1832
Db 1130 IDLLTNLIFANGISTWGDNTSISTTL---NVSSNGIQGVIAEDAQ---VNAITTGTTTI 1184
Qy 1833 TVSSDGIS-----IQCKNSSITLSKDLNMGVSKVTSNVGKPKDPTDAANV-QOL 1881
Db 1185 KIQDNANANFSQTAAYTLIQGGRFNGTLAGPNAFVGTSGNIFVKYELIRDSNQDYVLTRT 1244
Qy 1882 NEVRNLL--GLGNAGDNADGNQVNIADIKDPNCGSSNSTVIAKAGTVLGKGNNDTEK 1939
Db 1245 NDVLNVTTAVGNSAIAANPQVSNISRCLESTNTAAYNNMLLAK-----DPSD 1293
Qy 1940 LATGCVQGVKDGKNGANGDLSNVVVKTKQKSGKALLATYNAAGQTNVLTNNPAEADRI 1999
Db 1294 VATFVGATATDTSAAVTTVNLNDTKQTD-----LLS--NRLGTLRLSLN--AETSD-- 1341
Qy 2000 NEQGRIRFHVNDGNEQPVVQG---RNGIDSSASGKHVSVAIGFQAKADGEAAVAIGROTQA 2056
Db 1342 -VAGSATGAVSGDEAEVSYGVWAKPEFYNAEQDKKGGIAGYKARTTG---VVVGLDPLA 1397
Qy 2057 GNOQ---STAIGNAQATGDQSIAITGTVVTKHSGAIGDPTVKADNSYSVGNNNQFID 2113
Db 1398 SDNLMIAGAAIGITTKDIKHQYKKDKTDINGLSFSLYGSQOLVK---NFFAQN----- 1449
Qy 2114 ATQTDVFGVGNITVTESVALGSNSAISAGTHAGTQAKKSDG---TAGTGTITTAGATGT 2170
Db 1450 ---AIFTL--NKVKSQRYEFESGKMSKQIAAGNYDNMTFGNLIFGYDYNAMPNVL 1503
Qy 2171 VKGFAGQTAGVAVS-----VGASGNERIQNVAAGEVSATSTDAVNGSQLYKQTGIANA 2225
Db 1504 VTPMAGLSYLSKSNENYKGTGTIVANKRINSKFSRV-----DLIVGAKVAGSTVNIIDI 1558
Qy 2226 TNEIDHRTHQ--NEKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGQGAVALGSKL 2284
Db 1559 V--IPELHSFVHVHVKLNSQSML-----DQOTAPFISQP 1594
Qy 2285 SDNGQWFKINGSADTQGHVGAAGVAGPHF 2314
Db 1595 DRTAKTSYNIGLSANIKSKDAKMEYIGYDF 1624

RESULT 7
OMP_RICRI
ID OMPB_RICRI STANDARD: PRT; 1654 AA.
AC Q53047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
GN OMPB.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R;
RX MEDLINE=92167802; PubMed=1724278;
RA Gilmore R.D. Jr., Cleplak W. Jr., Policastro P.F., Hackstadt T.;
RT "The 120 kilodalton outer membrane protein (rOmp B) of Rickettsia
RT rickettsii is encoded by an unusually long open reading frame:
RL evidence for protein processing from a large precursor.";
RL Mol. Microbiol. 5:2361-2370(1991).
RN [2]
RP SEQUENCE OF 279-1654 FROM N.A.
RC STRAIN=R;
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
```

```
RT "Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii.";
RL Mol. Microbiol. 3:1579-1586(1989).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16353; CAA34403.1; -.
KW Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1 1333
FT CHAIN 1334 1654 32 KDA BETA PEPTIDE.
FT DOMAIN 1181 1188 POLY-THR.
FT SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;
SQ
Query Match 3.8%; Score 444.5; DB 1; Length 1654;
Best Local Similarity 21.8%; Pred. No. 5.6e-09;
Matches 409; Conservative 182; Mismatches 728; Indels 561; Gaps 88;
Qy 662 ITPAGENGISINDIAKGVKVGIDPINGL---TTPKLTGVSKDK---GKTQLVIEQVAS 715
Db 72 ITANANINGINFNPAGS-----FNGLLNTANNLAVTVSEDTTLGFTINNVHNAHS 122
Qy 716 GNDTKNIIRGLSPITLPSITNAGGVRTTBOGNTITSDKSKAASIGDILNTGFNLKNSN 775
Db 123 ENLTLNACKTLITIGQVYNAAQAATKNAQNVVQ-----FNGAIDNND- 168
Qy 776 SVGFVSYNTVDFIDGNATTAKVYDETQNTSKVYDVNVDEKTIELGSD-----NGKTN 830
Db 169 -----LKGVGRIIDFGAPASTLVFNLANPTT-----QKAPLIIGDNVANGVNG 212
Qy 831 KIGVKTITLTITN---ANGKATNFSTTDNDALVNAKDAENLNTLAKETHT---KCTA 883
Db 213 TLNTNGFIQVSNKSFATVKAINIA--DQGLIFNTD-ANNANTLNLOAGGTTINFTGTD 269
Qy 884 DTALQTEKVKKDGATDDETTITVGKDGTONK--TVNTLKLKGENGLTVATNKDGTVTFCIN 942
Db 270 GTG-RLVLLSKHAAATNFNITGSLGSLKGVIEFNTVAVDQG--LTANAGAANAV-IGTN 325
Qy 943 TQSLKAGDSTTLNKDGLSIRKPASNEOIQVADGVKFAKVD-----KGNSTGIDGTSR 997
Db 326 NCAGRAAG-----FVVSVDNGKVATIDGQVYAKDMVYIQSANATGQ 365
Qy 998 ITKQIGFTGANGSLDTPKPHLTOKLKVGEVITNTGINAGGKKTITNQSDITONSND 1057
Db 366 VNFRIHVDVGADGT-----TAFKTAASKVT-----ITQDSNF 397
Qy 1058 AVT--GGRVYDLKTELESKINSAAKTAQNSLHEFSVADEQGNHFT--VSNPYSSYDTSKT 1113
Db 398 GNTDFGNLAAQIKVP-----NAITLT-----GN-FTGDASN-----GNT 431
Qy 1114 SDVITF-----AGENGITTKVNGVVRVIGIDQTKGLTTPKLTGVGN-- 1155
Db 432 AGVITFDANGTLESASADANAVTNITAIIDASG---AGVVQLSGTHAAELRLGNAGSIF 488
Qy 1156 --KGIVIDSK-----DGONTITGLSNTLNTVNDGAGHLSQ--GLA 1194
Db 489 KLADGTVINGKVNQALVGGALAAAGTITLDGSATITG-----DIGNAGGAALQRIILA 542
Qy 1195 NDTDKTRAASIGDVLNAG---FNLOGNCEAVDFVSTYD--TVDFIDGNATTAKVYDDTS 1249
Db 543 NDAKKTTLTGGANITGAGGGTIDLOANGGTIKLTSTQNNIVVDFDLATATDQTGVVDASS 602
```



QY 1250 KTSKVVYDVN-----VDNKTIETVSDKKLGKVTTLTKTSANGNA--TKESAADGDAL 1300  
DB 603 LFNATLTINGKIGTANGKTL--GQFNIGSSKTVL-----SNGVAINELVINDGAV 655  
QY 1301 VRASDIATHLNTLAGDIOTAKGASQASSASVVDADGNKVIYDSTDKKYQVNDKGVDPK 1360  
DB 656 QFHD--TYLIT-----RITNAAGQ-----GKLIENPV-----VNN 684  
QY 1361 NKEVAKDLVAQAQTPDGTGLAQMNVKSVINKBOVNDANKKQGINEDNAFIKLENAAKDT 1420  
DB 685 GTTLAAGTNLGSATNP--LAEINFGS-----KGVNVD----- 714  
QY 1421 KTKNAAVTVGD--LNAVAQTPPLTFAGDTGTTAKKLGETLTIK---GCQDTNKLTDNNICV 1476  
DB 715 ----TVLNVGEGVNLATNITTTDANVGSVFVFNAGGTNIVSGTVGG-----QGNKFNT 764  
QY 1477 VAGTGGFTVKLAKDLTNLSNVNAGGTTRIDEKGISFVDANGQAKANTPVLNGLDLGGKR 1536  
DB 765 VALENGTIVKFL-----GNATFNGTITIAANSTLQIGGNY 799  
QY 1537 ISNIGAAVDDNDVAVNFKQFNEVAKTVNN-----LNNQNSG----- 1572  
DB 800 TADCVASADGTGIVEFVNTGPTITVLNKQAAPVNALKQITVSGPGNVVINEIGNAGNHG 859  
QY 1573 -----ASLFFVVDANGKPDNGDKPKQKAIKAGADKYHANANG--VP----- 1614  
DB 860 AVTDITAFENSILGAVFLPRGIPENDAGNTMPLTIKSTVG---NKTAKGFDVPSVVLVG 916  
QY 1615 ---VVDKGPITDADKLANLA-----AHGKPLDAGHQVVASLGGNSDAITLTIKST 1663  
DB 917 VDSVITADGQVIGQNNVGLGSGNGIILVNATTIYAG---ISTLNNNGTIVLSG----- 969  
QY 1664 LPQIDTPNTGNAGQAQSLPSLSAAQSVKVDLVNFGVNLQTNHNOVDVFK----- 1718  
DB 970 ---GVNPTGTVYGLG---TGIGASKFKQVTFDDTYNNLGNLIATINADGVIVTTGG 1022  
QY 1719 ---AYD---TVNFVNGTGADITSVRSADGTMSTNTVNTALATDDGNVLKAKDGKPY 1771  
DB 1023 IAGIGFDGKITLGSVNGNG---NVRFADGILSN---STSMIGTTKANNGTVYLGNAFV 1075  
QY 1772 K--ADDLMPGSLKAGKASDAKT-----PTGLSLVNPVNAKGSTGDAVALNN 1817  
DB 1076 GNIGSDTPVASVTRTGSAGLQNIYSQVIDPCTYNLGIVNSNIILG---GGTIAING 1133  
QY 1818 LSKAVFKSKDGTITTVSSDGSIGQKDNSSI--TLSKDCLNVGKVI---SNVKGKTKD 1872  
DB 1134 -----KIDLVNTLTFSAGTSTWG--NNTSIETTLTLANGNIGHIVILEGAQVNTTTG 1185  
QY 1873 TDAANVQQLNEVRNLLIGLGNAGDNADGNQ-----VNIADIKDPN---SGSSS--N 1919  
DB 1186 TTTIKVQD-----NA--NANFSGTQTYTLIQGGARFNGTLGSPNFVAVTGSNRFVN 1233  
QY 1920 RVVIRAGTVLGGKGNNDTEKLTATGGVQGVQVDK-----GNANGDLSNV--WVKTKDGS 1971  
DB 1234 YSLIRA-----ANQDYVITRTNAENVVTVNDIANSFPGFAPGVQDVNTVFNATNTRAA 1286  
QY 1972 KKALLATYNAAGQTNY---LITNPAEADIRNEQIRFPHVNDGNQ----- 2014  
DB 1287 YNNLLAKNSANSANFVGAIVTDSAAITNVQDLDAKDIOAQLGNRGLALRYLGTPEATA 1346  
QY 2015 -----EPVVOGRNIDSSASG-----KHSVATGFOKADGEARAVALIG 2051  
DB 1347 MAGPEAGAIISAAVAAGDEAIDNVATGVIAKPFYTTDAHQSKKGLAGYKAKTTG---VVIG 1403  
QY 2052 ROTOAG-----NOSIAGDNAQATGDQGSIAIGTGNVVTG--KHSGAIG 2092  
DB 1404 LOTLANDNLMIGAATGITWTDIKHODYKGDKTVDNVSFSLYGAQQLVKNFPAOGSAIF 1463  
QY 2093 DPFSTKADNSY-----SVGNNNOFIDATQDVFVGNGNITVTSNVALGNSAISAGTHA 2148  
DB 1464 SLNQVKNKSQRYFFDANGNWSQIAAGHYDNMTFGNLTVGVYDYNAMQGVLPVPMAGL-- 1521  
QY 2149 GTQAKKSDGTAGTITTTAGATGTGVKG--FAGQTAVGAVSVGASGAERRIIONVAAGEVSATST 2207

DB 1522 -SYLKSSDSENYKETGTTTAVANKOVNSKFSQRTDL-IVCAKAVAGSTMNITDLAV-----Y 1572  
QY 2208 DAVNSQLKATQGIATNATNELDHRIT-----HONENKANAGISSAMAMASMPQAYIP 2259  
DB 1573 PEVHAFVHVHKVTCRLSKTOSVLDGQVTPCINQPDRTTKTSYNLGILSASIRSADKMEYGIG 1632  
QY 2260 GRSMVTGGIATINGOGAVAV 2279  
DB 1633 YDAQISSKYTAHQGTCLKVRV 1652  
RESULT 8  
OMPB\_RICTY  
ID OMPB\_RICTY STANDARD; PRT; 1645 AA.  
AC P96989;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
GN SLP.  
OS Rickettsia typhi.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=785;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WILMINGTON;  
RX MEDLINE=94040787; PubMed=8224886;  
RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;  
RT "Cloning and sequence analysis of the gene encoding the crystalline  
RT surface layer protein of Rickettsia typhi.";  
RL Gene 133:129-133(1993).  
RN [2]  
RP PARTIAL SEQUENCE.  
RC STRAIN=WILMINGTON;  
RX MEDLINE=92114896; PubMed=1370573;  
RA Ching W.M., Carl M., Dasch G.A.;  
RT "Mapping of monoclonal antibody binding sites on CNBR fragments of  
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia  
RT prowazekii.";  
RL Mol. Immunol. 29:95-105(1992).  
RN [3]  
RP IDENTIFICATION OF CLEAVAGE SITE.  
RX MEDLINE=92104668; PubMed=1729180;  
RA Hackstadt T., Messer R., Cieplak W., Peacock M.G.;  
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer  
RT membrane protein of rickettsiae: Identification of an avirulent  
RT mutant deficient in processing.";  
RL Infect. Immun. 60:159-165(1992).  
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.  
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A  
CC S-LAYER WITH HEXAGONAL SYMMETRY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L04661; AAB48987.1; -  
KW Antigen; S-layer; Transmembrane; Cell wall.  
FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.  
FT CHAIN 1354 1645 32 KDA BETA PEPTIDE.  
FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).

Db	821	DPITVTLLNKQAYFCVLKQVMVSGPCNIIAFNEIG-----NGVAHA-----	860
Qy	1535	KRISNICAAAYDDDAVNFKQFNEVAKTVNNLNQNSGASLPFVYTTDANGKPIGTGDKP	1594
Db	861	-----IAV-----DSISFE-----NASLGNLSL-FLI-----SGPLD-----	886
Qy	1595	OKAIKGADGKYYHANANGVPVDKDKGPITDADKLANLAARHKPLDAGHQVVASLGNSDA	1654
Db	887	VLTIKSTVG-----NGT-VDNFNAPILVSGIDSMINQGVIGDQKNIIALSLGSDNS	938
Qy	1655	ITLJNIKSTLPQIDPTNPGNANAGQAQSPSLSSAAQOSNAASVKDVLNVGNFLQTH---	1711
Db	939	ITV-NSNTLYAGIRTTKTN-----QGTVTLSGGIPNNPGTII-----YGLGENDGPKL	985
Qy	1712	NOVDFVKAYDTVFVNGTGDADITSVSADGTSNITVNT-ALAAATDDDGNNVLIRAKDGF	1770
Db	986	KQVTFTTDYANLGSIIATNVIN-----DDVLTITGGIAGTDFDGKITLGSING--	1034
Qy	1771	YKADDLMPGSLK-AGKSASDAKTPTGLSLVNPAGKGTGDAVALNNLSKAVPKSKDGT	1829
Db	1035	-----NANVKFVDRTFESH-----PTSMIV-----ST-----KANQGT	1061
Qy	1830	TTTTVSSDGIIOGKNSSITLSKQGLNVGKVISNVGKGTGKTDDAANVQQLNEVRNLLG	1889
Db	1062	VT-----YILGNALVGNI-----GSSDIPVASVR-----	1084
Qy	1890	LGNAGNDADGNQVNIADIKKDKPNSGSSSNRTVIKAGTVLGGKGNNDTEKLGQVGVV	1949
Db	1085	--FTGNDSGVGLQGNTHSQNID--FGTYNLTLILSDVILGG-----GTTAINGEI	1130
Qy	1950	DKDGN-----ANGDLSNVWVKTKDQGSKKALATYNAAGQTNLYTNPAEADIRNEOGIR	2005
Db	1131	DLTLNLIFANG-TSTW-----GNNTSLSTLN-----	1157
Qy	2006	FFHVNDGNQPVVQGRNGIDSSASGRKHSVAIGFQAKADGEAAVAIGROTQAGNSIAIGD	2065
Db	1158	-----VSGNVGQIV-----IAEGAQVNA-----TTTGTTTIKIQD	1188
Qy	2066	NAQA--TGDSQIAIGTGNVTVGKHSICAIGDPSTVRADNSYSYVGNNNQFI-----	2117
Db	1189	NANANFGTQTYTILQGG--ARFNFTGLAP-----NFDVTGNIFVKVELIRDANQD	1238
Qy	2118	DVFCVGNNITVESNSVALGNSAI--SAGTHAGTQAKKSDGCTAGTTTTAGATGTVKGFA	2175
Db	1239	YVLTRTDVLNVVTTAV--GNSAIANAPGVHONTAI-----CLESTDTAAYNNML--LA	1288
Qy	2176	QGTAVGAVSVGSGAERRQNVAAQEVSTATSDVANGSOLYKATOGIANATNEDLRIHQ	2235
Db	1289	KDSSDVATFICA-----IATDTGAAVATVNLNDQT--KTQDLLGNRLGAL--RYLS	1335
Qy	2236	NENKANAGISSANAMASMPQA-----YIPGRSMVTGGIATHNGOGA-VAVGL	2281
Db	1336	NSETADVGGSETGAVSGGDEAIDQVSYGVWAKPFYINAEQDKKGLAGYKAKTAGVVVGL	1395
Qy	2282	SKLSDNGQWVFKINGSADTQGHVGAAGV	2309
Db	1396	DTLANDNL-----IGRAIG	1410
RESULT 9			
P2P_LACPA			
ID	P2P_LACPA	STANDARD;	PRT; 1902 AA.
AC	Q02470;		
DT	01-JUL-1993 (Rel. 26, Created)		
DT	01-JUL-1993 (Rel. 26, Last sequence update)		
DT	15-JUL-1998 (Rel. 36, Last annotation update)		
DE	PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-		
DE	ASSOCIATED SERINE PROTEINASE) (LP151).		
GN	PRTP.		
OS	Lactobacillus paracasei.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;		
OC	Lactobacillus.		
OX	NCBI_TaxID=1597;		

RN SEQUENCE FROM N.A.  
 RC STRAIN=NCDO 151;  
 RX MEDLINE=92381481; PubMed=1512565;  
 RA Holck A., Naes H.;  
 RT "Cloning, sequencing and expression of the gene encoding the cell-  
 RT envelope-associated proteinase from *Lactobacillus paracasei* subsp.  
 RT paracasei NCDO 151.";  
 RL J. Gen. Microbiol. 138:1353-1364(1992).  
 RN [2]  
 RP SEQUENCE OF 189-196.  
 RX MEDLINE=92226694; PubMed=1564442;  
 RA Naes H., Nissen-Meyer J.;  
 RT "Purification and N-terminal amino acid sequence determination of the  
 RT cell-wall-bound proteinase from *Lactobacillus paracasei* subsp.  
 RT paracasei.";  
 RL J. Gen. Microbiol. 138:313-318(1992).  
 CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
 CC -!- GROWTH OF THE BACTERIA ON MILK.  
 CC -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD  
 CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,  
 CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND  
 CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS.  
 CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED  
 CC INSULIN B-CHAIN.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
 CC SUBTILASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; M83946; AAA25248.1; -;  
 DR PIR; B44858; B44858.  
 DR HSSP; Q99405; LMPT.  
 DR MEROPS; S08.019; -;  
 DR InterPro; IPR000209; -;  
 DR InterPro; IPR001899; -;  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam; PF00082; Peptidase\_S8; 3.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; FALSE\_NEG.  
 KW Hydrolase; Serine protease; Cell wall; Zymogen; signal;  
 KW Transmembrane.  
 FT SIGNAL 1 33 POTENTIAL.  
 FT PROPEP 34 187 POTENTIAL.  
 FT CHAIN 188 1902 PII-TYPE PROTEINASE.  
 FT DOMAIN 188 1876 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1877 1895 MEMBRANE ANCHOR (BY SIMILARITY).  
 FT DOMAIN 1896 1902 CYTOPLASMIC (POTENTIAL).  
 FT ACT\_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE  
 FT PROTEINS.  
 FT SEQUENCE 1902 AA; 200253 MW; D8C9F38CEE5DA582 CRC64;  
 Query Match 3.5%; Score 407.5; DB 1; Length 1902;  
 Best Local Similarity 21.1%; Pred. No. 1.4e-07;  
 Matches 450; Conservative 262; Mismatches 856; Indels 567; Gaps 109;  
 QY 400 GLKVKLAKN--LSGLETVSTKNLTA---SEKVTYVSGNNTAEQLQSGGLTFTPTTNASTD 453  
 DB 7 GLSILAGTVAGALAVLPVGEIQAKAASISQPKVSSLANTVRAAKAQAATDTTAATTN 66

QY 454 KTVYGTDLKFTDNSNTALEDDTTRITKDKIGFSNKAGTYVDENKPYLDKDKLVGNSTLNN 513  
 DB 67 QAI-----ATQLAAKGIDY-NKLNVKVOQDITYVDVIVQMSAAPASEN 107  
 QY 514 GGLTVNNTGGSNKQIQVGADGKIFADVNVNVNNAKFGCTTRITEEIGFADADGKV--- 570  
 DB 108 GTL---RTDYSSTAETIQETNKVIAQAQSV-----KAAVEQVTQQTAG--ESYGYVNG 156  
 QY 571 -DKKSPYLDKKQL-QVGGVKIT-----KDSGINA-----GQOKISNVK 606  
 DB 157 FSTKVRVVDIPKIKQIAGVKTYTLAKVYPTDAKMSMANVQAVWSNYKYKGEGTVSVSI 216  
 QY 607 DATDDTDAVYIKOLKQVQDADGALQSFSIRDEKGOEFTISNLYSNGNTPTTETITFAG 666  
 DB 217 DTGIDP---THKDMR-LSDDKDVKLTKYDV--EKFTDTAKHGRYFTSKVPYGEN---YA 266  
 QY 667 ENGISISNDIAKVKVGVGIDPINGLITPKLTVGSDCKGTQLVIEQVAVSNDYKNIIRGL 726  
 DB 267 DNNDTITDDT-----VDEQGHMVGAGI-IGAN-----GTGDDPTKSYVG 305  
 QY 727 SP-----TLPSITNAGGVRTTEQNTITTSDEKSKAASIGDIILNTGPNLKNNSN----- 776  
 DB 306 APAQALLAMKVFTNSDTSATTSATLVSALSDSAKIGA--DVLNMSLGSDSGNOTLEDPE 363  
 QY 777 VGFVSTYNTVDFIDGNATTAKVTVYDETNTQTSKYTVYDYNVDEKTIELTGDKNGKTNKIGVKT 836  
 DB 364 IAAVQANNE-----SCTAAVISAGNSGTSGSATQGVNKDYGLQ---DNEMVGTGPG--T 412  
 QY 837 TTTTNTNANGKATNFSTDDNDALVNAKDIENLTLAKEIHTTKGTADTALQTFKVKKD- 895  
 DB 413 SRGATTVASAEITDV-ISAQVTTIDGKQLQGPETITQLSSNDEFTGSPDQ--KKFYVVKDA 469  
 QY 896 -----GATDEE-----ITVGKDGTON-GKVTNTLLKGENGLTVATNKDGTV----- 937  
 DB 470 SGDLSKGAADYTDADAKGTAIVKRGELNFPADKQKVAQAAGAGL-IIVNDGTATPLTS 528  
 QY 938 -----TFGINTQSGLKAGDSTTLN-KDGLSIK-----NPA 966  
 DB 529 IRLTTTPTFGLSKTKQKLVDMVTAHPDDSLGVKIALTLPLNQKYTEDKMSDFTSYGPV 588  
 QY 967 SNEQIQ--VCADGVKFAKVDKGNSSGIDGTSRTIKDQIGFTGANGSL----- 1012  
 DB 589 SNLSFKPDITAPGGNIWSTQNNNGYTNMSTSMASP-----FIAGSQALLKQALNNKNPF 644  
 QY 1013 -----DTTKPHLTDKLKVGEVEITNTGINAGGKKITNIQSGDITQNSNDAVT-----GGR 1063  
 DB 645 YADYKQLKGTALTDFLKTVMEN-TAQPIN-----DI--NYNVIVSPRQAG 689  
 QY 1064 VYDLKTELES-KINSAAKTAQN-----SLHEFSVADBEQ-----GNHFTVSNPYSSYDTSK 1112  
 DB 690 LVDVKAIDAILEKNPSTVVAENGYPVAVELKDFSTDKTKFLTNTNRTTHELTQYQMSNTD 749  
 QY 1113 TSDVITFAGE--NGI---TTKVNKGVVRVIGIDQT----- 1141  
 DB 750 TNAVYTSATDPNSCVLYDKKIDGAAIKAGSDITVPAGKTAQIEFTLSLPKSFQDQOQFVEG 809  
 QY 1142 ----KGLTTPKLTIV-----GN-NNGKGIYDSDKG-QNTITGLSNTLANVTNDGAGHA 1188  
 DB 810 FLNFKSDGSRNLNLPYMGFFGDWMDGK--IVDSLUNGITYSPAGGNYCTVPLLTNKNTHQ 867  
 QY 1189 LSGQLANDTDTKRAASIGDVLNAGFNLCQNGEAVDFVSTYDVTDFIDGNATTAKVTVYDDT 1248  
 DB 868 YGGMVTADGKQTV-----DDQATAFSS-----DKNA-----LYNDI 900  
 QY 1249 SKTSKVYVDVNVNKNKTEIETVSDKKLGKVTTL-----TKTSANGNATKP-----SAADG 1297  
 DB 901 SMOYLLR--NISNVQVDIILDGQ--GNKVTTLSSSTNQTKTYDDAHSOKYIYTNAPAWDG 956  
 QY 1298 DALVKASDIATHLNTLAGDIQTAKGASQASSASYVDADGNKVIYD---STDKKYQV-- 1352  
 DB 957 -----TYYDQROGNIKTADGSYTYRISGVPGGDKQVDFVPFKLDSKAPTVRH 1006

```
QY 1353 -----NDKGOVDKNKEVAKDLVLAQAQTPDGTFLAQMNVKSVINKBOVDNANKKQGIN 1404
Db 1007 VALSAKTEGKQYYLTAE-AKDDL-----SGLDATKSVKTAINEVTNLDATFTDAGT 1058
QY 1405 EDNAFTKGLNAAKDTKTNAATVTVGDLNAVATPLTFTAGDTGT-----TAKKLGET---L 1457
Db 1059 TADGYTK-IETPLSDEQ-----AALGNGDNSAELVLTDNASNATNODASVQKPGSTFSL 1113
QY 1458 TKGGQTDNKLTDNNIGVAGTDG-----FTVKLAKDLTNLNSVNAAGSTRIDEKGIS 1510
Db 1114 IVNGGGI-PDKISSTTGYEANTQGGTYTFSGTYPAAVD-----GTYTDAQGGK 1162
QY 1511 FVDANGQAKANTPVLSANGLDLGGKRISNIGAAVDDNDVAFNKQFNE-----VAKTVNNL- 1565
Db 1163 H-DLNTTDAATNTSTASMAVNTADYAQVDLYADKAHTQLLKHFDTKVRLTAPTFTDLK 1221
QY 1566 -NNQNSGASLFFVYT---DANGKPIGTGDKPKAIGKADGKYYHANANGVPV----- 1615
Db 1222 FNGSDQTSKATIKVTGTVSSDTKTVNVD-----TVAALDAQ---HHFSVDVPVNYGDN 1274
QY 1616 -----DKDG-----KPIIT---DADKLANLAHKGKPLDAGHVVASLGNSDAITLNI 1660
Db 1275 IKVTATDEGNTTEBQKTISSYDPLKNAV-----FDQG-----VKFGAN-----EF 1319
QY 1661 KSTLPQIDTPNTGNANAGAOQSLPSAAQOSNAASVKDVLNVGPNLQTNH-NQVDF-VK 1718
Db 1320 NATSAKFDPKGTGATITGKVKHPTTTLQVQDKQISIKNDLTFSTLDTGLTGKPPFGV 1379
QY 1719 AYDVTNVFNGTGAD-ITSVRSADGTMSTNVTNLTAAATDDGDNVLI---KAKDGKPKADD 1775
Db 1380 VGDITQ--NKTFOEALFTILDVAFTLSLSDSTDPVYTNDFEITGTATDQAQYLS-- 1435
QY 1776 LMPNGSLKAGKSADAKTPTGLSLVNPNGKG---STGDVAL-----NNLSKAVFKSKDG 1828
Db 1436 LAINGSHVASQAD-----INISGKPGHMAIDQPVKLLGKKNVLTAVATDSENN 1485
QY 1829 TTTTIVSSDGISIOGKDN-----SSITLSKDLNGLVGGKV--ISNVGKGTKD 1872
Db 1486 TTTTKIT---VYIEPKKTLAFTVTPSTTEPAKVTTLTANAAATGETVOYADGGKTYQD 1542
QY 1873 TDAANVQ-QLNEVRNLLGLGNAGDNADGNVNIAIRKDP-----NSGSSSNRTVIKAGT 1927
Db 1543 VPAAGVTVTANGTFKFKSTDLVGNESPAVDYV-VTNIRKADDPALQOTAKAALTNLIASAK 1601
QY 1928 VLGGKGNND-----TEKLTAG--GVQGVQDKGNANGDLSNVVWKTQKDGSKKAL 1975
Db 1602 TLSASGKYDDATTTALAAATQAKQALDQTDASVDSLGTGANRDLQT-----AINQL 1652
QY 1976 LATYNAAGOTNYLTNNPAEADIRNEQIRFFHVNDGNQEPVQGRNGIDSSASKHVA 2035
Db 1653 AKLPADKKTSL-----NQLOSVKAAALGTDLGNQTDPSGTG 1688
QY 2036 IGFOAKADGEAAVATGROTAQAGNQSIAIGDNAQATGDSIAIGTGNVVTGKHSIGAIGDPS 2095
Db 1689 KTFEAAALDLVA-----QAQAGTQT---ADQLQA-----SLAKVLDLAVLAKLAEG- 1730
QY 2096 TVKADNSVSVGNNOFIDATQTDVFGVGNNTVTVESNVALGSN-----SAIS 2143
Db 1731 -IKAATPAEVGNAK-----DAATGKTYADIAIDLTSQGSASADSKLAHLQALQSLKTKVA 1786
QY 2144 ACTHAGTCAKSGDGTAGTITTTAGATGTVKGFAGOTAVCAVSVG---ASGAERRIONVAAG 2200
Db 1787 AAVEAAKTAGKDDTTGTSGDKGGGTGTPAPAGDTGDKDGEGSQPSGG-----NIPTK 1841
QY 2201 EVSATST---DAVNGSOLYKATQGIANNATNLDHR 2232
Db 1842 PATTTSTSTDDTTDRNGQHTSGKALPKTAETTER 1876

RESULT 10
CBPA_CLOCL
ID CBPA_CLOCL
AC P38058;
```

```
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CELLULOSE BINDING PROTEIN A PRECURSOR.
GN CBPA.
OS Clostridium cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
RN NCBI_TaxID=1493;
RP [1]
RX MEDLINE-9222810; PubMed-1565642;
RA Shoseyov O., Takagi M., Goldstein M.A., Doi R.H.;
RT "Primary sequence analysis of Clostridium cellulovorans cellulose
binding protein A.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3483-3487(1992).
CC -|- FUNCTION: BINDS TO CELLULOSE FIBERS AND COORDINATES CELLULOSE
ENZYMES.
CC -|- PTM: THE N-TERMINUS IS BLOCKED.
CC -|- PTM: CONTAINS ABOUT 10% OF CARBOHYDRATES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
EMBL: M73817; AAA23218.1;
PIR: A44140; A44140.
HSP: O06851; INBC.
InterPro: IPR001956;
InterPro: IPR002102;
Pfam: PF00942; CBD.3; 1.
Pfam: PF00963; Cohesin; 9.
KW Cellulose degradation; Cell wall; Repeat; Signal; Glycoprotein.
FT SIGNAL 1 28
FT CHAIN 29 1848
FT DOMAIN 29 189
FT SEQUENCE 1848 AA; 189152 MW; 85FA6CE6F771AF1A CRC64;
SQ
```

Query Match 3.5%; Score 406; DB 1; Length 1848;  
Best Local Similarity 20.6%; Pred. No. 1.5e-07;

Matches 452; Conservative 277; Mismatches 800; Indels 668; Gaps 111;

```
QY 31 SSSTAGQVGS-SPVRLTRVATLALV-----ICATLNGSAY 66
Db 39 NSNKAQNTSITPIIKITNTSDSLNLDVKNVRYVYSDGTQGTFCWDHAGALL-GNSY 97
QY 67 AONNSKIAFGTTGNNDNASNEASIAIG---SLAKAHANQAIAGG---SKPDPNR----- 117
Db 98 VDNISKVTANFVKETASPTSTYDYVEGFGAGRATLKKGFITIQGRITKSDSNYQT 157
QY 118 -----QAANQKAGSHAKKESI--AIGDVLAEQASTAIGSDDLYLDRNSTNS 164
Db 158 NDYSFDASSSTPVYVNVKVTYIGGAKVLGTAPGPDV-----PSSIINPTSATFOKNVTQ 212
QY 165 KYPNGLLSTLIQNTVLRQIRDSNGSQKRRTAEGHASTAVGA--MAYAKGHEANFGT 222
Db 213 A---DVVTFMTLNGNTFKTITDANG-----TALNASTDYSVSNDVTLISKAYLAKOSVG 263
QY 223 RSTAEKNYSL-----AVGLT---AKAEKGYTIAIGSNAQAINTYCALAL- 262
Db 264 TTTLNFESAGNPQKLVITVVVDTPVEAVTATIGKVQVNAGETVAVPNLTKVPAAGLATI 323
QY 263 -----GADTRVDLDYGIALGYSQILNNNNNNNNNNKAYVEPEGNSNIK-----SS 306
Db 324 ELPTFDSASLEV-----VSTAGDIVLNPVNSSTV-----SGSTIKLLFLDLDLGSQ 373
QY 307 KATGNGLFSGSSTIKRKIIN-----VGAGYEDTDVAVNAQLK-----AVENLAKRQITF 356
```

Db 374 LITKDGVEA--TITPKAKAITGTTAKVTSVKLAGTVPVGDALQOEKPCAVN---PGTVTI 428  
QY 357 KGDNDGTGVKKKLGELTITIKGGETOADKLTDDNNNIGVWTDNNNTGLKVKLAKNLGLSETYS 416  
Db 429 NPIDNRMQI--SVG-TATVKAIEAAPPV---LTSVPSTG-----IA 465  
QY 417 TKNLTAASEKVTVCSGNNTAELOSGGLTFPTTNASDTKTVYGTGDLKFTDNSNTA-----L 472  
Db 466 TABAQVSFDATL---LEVASVTAGDIVLAPPTVNFSS---YTNGVNIKLLFL 510  
QY 473 EDT---TRITKD---KIGFSNKAAGTVDENKPYLDKDKLKVGNSTLNNGLTIVNNTIGGS 525  
Db 511 DOTLGSOLSKDGVFTVTFKAKAVSTVTP-----VTVSGT----- 548  
QY 526 NKQIQVAGADIKFADVNVNVSNAKFGTTRITEE--EIGFADADGVKKSPYLDKQLQ 583  
Db 549 ----PVFADGTL---AEVQSKTAAGSVTINIGDPILEPTISPTATFDKAPADVATTMT 601  
QY 584 VGGVKTIKDSGINAGDOKIS-NVKDATDODTAVYKQLKQVQDADGALQSFIRBEKQ 642  
Db 602 LNSYTFNGITGLTSDYSISGNVVKIS-----QAYLAKQPVG- 638  
QY 643 EFTISLNSNGNTPNFTFITPAGENGISINDIAKGVKVG-----IDPINGLITP 694  
Db 639 DLTLTFNSGNKTATAKLVSSIKDAPKTVTATGTATVNAAGTAVPVTLNSVSGISTA 698  
QY 695 KLVGSDKDGKTOLVIEQVAGSNDTKNIIRGLSPITLPSITNAGGVRTTEQGNITSDDEK 754  
Db 699 ELQLSFD---ATLEVVISITAG---DIVLNPVSFNVSVN-----GSTI----- 736  
QY 755 SKAASIGDILNCGFNKLNNSNSVGFSTVNTVDYFDIGNATTAKVTVDEINQSKVITYDVN 814  
Db 737 -KLLFDDDTLGSOLSKD-----GVFATINF-RAKSVTSTVTPVKVSGTPPFADGT 786  
QY 815 VDEKTIETLGDNGKTKKIG-VKTTILITTTNA---NGKATNFSTDDNALVNAKDAENLN 870  
Db 787 LAELSVETVAGSVTINAIGPVKTVTATGTATVKSETVAVPV---LSNVPCIA---- 838  
QY 871 TLAKETHTTKGADTALQFVKKKGDATDETTIVGK-----DGTONGKTVNTLKL 921  
Db 839 -----TAELQLSFDATLEVA---SITVGDIVLNPVSFNVSVNGSTIKLLFL 883  
QY 922 KGENGLTVATNKDG---TVTFGINTOSGLKAGDSTTLNKGSLIKNPASNEQIOVGADGV 978  
Db 884 DDTPLG-SOLISKDGLVATINFKAKTVS-----TVTPVAVSGTPPFADGT 938  
QY 979 KFAKVDKGSSTGIDGTSRTIKDQIGFTGANGSLDTTKPHLTDKDLKVG-----VEITN 1033  
Db 929 -LAEL---QSKTVAGSVTIEPSQ-----PVKTVTATGTATVKSETVAVPVTLN 975  
QY 1034 -----TGINAGKKITNTQSGDI-----TQNSNDVATGGRVY-----DLKTELES 1073  
Db 976 VPGIATAELQVGFADATLLEVASITVGDIVLNPVSFNVSVNGSTIKLLFDDTLGSLIS 1035  
QY 1074 K-----INSAKTAQNSLHEFSVADEQGNHFTVSNPYSSYDTSKTSVDITFAG-----EN 1123  
Db 1036 KQGLVATINFKAKTV-----TSKVTTPVAVSGTPPFAD 1068  
QY 1124 GITTKVNGKVW--RVGIDOTKGLTTPKLTGVNNNGKGIVIDSKDQNTITGLSLANVT 1181  
Db 1069 GTLAEINMKTVAGSVTIEPSQPKVTVATVG-----TATVKSETVAVPV--TILSNVP 1119  
QY 1182 NDGAGHALSQGLANDTKTRAASIGD-VLNAGFNLOG--NGEAVDFVSTYDVT--DFIDG 1236  
Db 1120 GIATAE-LQVGFDATLLEVASITVGDIVLNPVSFNVSVNGSTIKLLFDDTLGSLISK 1178  
QY 1237 NATTAKVTYDDTSKTSKVYDVNDVNDKI-----EVTSDKGLGVKVTTL 1280  
Db 1179 DQVLATINFKAKTVTSKVTTPVAVSGTPPFADGTALAEKTVETVAGSVTIEPSQPKVTVTA 1238  
QY 1281 TKTSANGNATKFSAAD-----GDALVKASDIATHL---NTLAGDITOTAKGASOAS 1328  
Db 1239 TVGTATGKVGETVAVPVTLNSNPVGIATAEVEQVGFADTLLEVASITAGDIV----LNPSVN 1294

QY 1329 SASYVDADGNKVIYDSTDKKYQVNDKGQVKNKEVAKDKLVAQAQTPDGTLLQAMNVKSV 1388  
Db 1295 FSSWNGSTIKILF-----LDDTLGSLISKDGVFTATINFKI- 1331  
QY 1389 INKEQVNDANKKOGINEDNAFIKG-LENAAKDTKTNAAVTVGDLNAYAVOPLTFAGDGTG 1447  
Db 1332 --KAVPSTGTPPAISGTPVFADGTAEVOYKTVAGSVTIAAADIKAVKAT-----VG 1382  
QY 1448 TTAKKILGETLITIKGGGTDTKNKLDDNNIGVAGTGDGFTVKLAKDLTNLN--SVNAGGTRID 1505  
Db 1383 TATGKAGDTAV-----PVTLSNVSGIATVE---LQLSFDATLLEVASITAGDIVLN 1431  
QY 1506 EKGISFVDANGQAKANTPVLSANGLD--LGGKRISNTIGAADDNDVAVNFQKOFNEVAKTVN 1563  
Db 1432 P-----SVNFSSVWNGSTIKILFDDTLGSLISKDGVFA---TVNFK---VKSTAT 1477  
QY 1564 N-----LNNQNSNG--ASLP-FVVTVDANGRPINTDOKPKQA-- 1597  
Db 1478 NSAVTPTVTVSGTPVFADGTTLAEELKSESAAGRLTILPTVILVDSTVAPTAVTFOKANOADA 1537  
QY 1598 --IKGADGKYVHANANGVPVDKDKPITTDADKLANLAHKGPLDAGHVVASLGGNSDAI 1655  
Db 1538 AITWTLNGNTFSAIKNGTATLVKGTDTVTSENVVTISK-----AYLAKQGTGTV 1585  
QY 1656 TL-----TNIKSTLPQIDTPTGNANAGQAQSLPSLSAAQQAQSNAAQSVKDVNLVGFN 1706  
Db 1586 TLEFVFDKGSNAKVVAVAKIEQIVNS-----TITPVVATFEKTAQAQADV---VT 1633  
QY 1707 LQTNHNOVDFVKAYDVNFVNGTGCADITSVRSADGTM--SNITVNTALATATDDGCVNLK 1764  
Db 1634 MSUNGNTFSAIK-----NGT---TTLVKGTDTYISGTVTISKAYLATLADGASATLE 1682  
QY 1765 --AKDKGFYK-----ADDLMPNGSLKAGK---SASDAKTPTGLSLVNPNAAGKST 1809  
Db 1683 FVFNQOASAKRLTITVPAVDPVPTDFAVKIDKVSAAAGSTVKVP--VSLINVS----- 1734  
QY 1810 GDAVALNNISKAVFK-SKQGTMTTTV--SDGISIQGKDNSSITLSKDGVLNGVKVISNVG 1867  
Db 1735 ---KGVNVCVAEYKISFSDSVLTVGTGTAGTSIK---NPAVNFS----- 1772  
QY 1868 KGTKDDTAAVQOLNEVRNLLGLGNAGN-----DNADGNQVNIAD-----IKKDPNSG 1915  
Db 1773 -----SOLN-----GNTITLFFDNTIGNELITADGQFATIEFKVNAA 1810  
QY 1916 SSSNRTV-IKAGTVLGGKGNNDTE--KLAT--GGVQV 1947  
Db 1811 ATSGTTAEVKVATISSFADASLATEITKVAIVNGSVKV 1847

## RESULT 11

HLVA\_SERMA  
ID HLVA\_SERMA STANDARD; PRT; 1608 AA.  
AC P15320;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-NOV-1990 (Rel. 16, Last annotation update)  
DE HEMOLYSIN PRECURSOR.  
GN SHLA.  
OS Serratia marcescens.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Serratia.  
OC NCBI\_TaxID=615;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.  
RC STRAIN=SN8;  
RA MEDLINE=88257037; PubMed=3290200;  
RX Poole K., Schibbel E., Braun V.;  
RT "Molecular characterization of the hemolysin determinant of Serratia  
marcescens";  
RL J. Bacteriol. 170:3177-3188(1988).  
CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD  
CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY







QY 716 GNDTKNIIRGLSPITLPSITNAGGVRIT-----EQNTIITSDEKSKAASIGDI 763  
Db 566 GKGDAEL--ALDPS-PVYKNSGGSYATSKNYIGVSAIFDQGEAGAPAKVPIPNVGA 622  
QY 764 LNTGFNLKNNSNGVFS-TYNTVDFIDGNATTAKV-----TYDENQTSKVT--- 810  
Db 623 QAPSAKGVNNGNATGYDLSHAKV---SGATGYKVQVFNKGFGFETLDLGNQTSWTTKG 678  
QY 811 ----TDVNVDEKTIELTDGNGKINKIGVKTTLTTTNNANGKATNFS-----TTDND 858  
Db 679 KIMPTSAEIRAGKVALHLKDGSGAELPIN-PGPTYKNAGDGAARVSKFIAYNKDGEA 737  
QY 859 LVN-----AKDAENLNTLAKELHTTKGTADTALQTFKVKDGATDDETIIVGKDGTQ 912  
Db 738 IASPAATPALPDIAKPNVTGYLYNTK-----SSQT 769  
QY 913 KTVNTLKLKGENGLTVATNKDGTVTFGINQSGLKAGDS---TTLNKDGLSKINPASNE 969  
Db 770 G-YVNLIIWEKVQNAKGYKVN---IYNGKEYQS-FDVGDAHWTTONKNLW---PTSEE 819  
QY 970 QIQVGADGVFAKVDKG-----NSSTGIDGTSRITKQDQIGFT-----GANGSLD 1013  
Db 820 ---IKAGSYKLHTDGGKGLALDPSPVYNNANGNYRG-----KKNYSFTLVAYDANGETI 871  
QY 1014 TTKP-----HLTKDKLKVGE---VEITNTGINAGKKITNTQSGDITQNSDNDVATGGRV 1064  
Db 872 PTAPFNPTFHEGAFFLTGEYWSLIIDPSQLNGA-----TGNVIYNEEDLSIDGRG 923  
QY 1065 YDLKTELESKINSAAKTAQNSLH-----EFSVADEQGNHFTVSNPYSSYDTSKTS 1114  
Db 924 PGL-----GLSRYSNLSOSSDLHFGQGWADAETSIVSTDGAMKIDEDATTHRETKA 977  
QY 1115 DVITEAGENCI---TKVKNKGVVRGIDQT-----KGLTPKLTGVNNGKGVIDSK 1164  
Db 978 DG-TYOPPGVYLELETAQDFLTATK--DOTNAYFNKKGKGLQKVVYDGHNNATVYTYNDK 1035  
QY 1165 DQNTITGLSNTLIANTDNCAGHALS--QGLANDTDKTRAASIGDVLNAGFNLQNGEAVD 1223  
Db 1036 NOLTAITDASGRKLTFTYDENGHVTSITGPKN--KKVTSYENDLLKKVTDGTVTSD 1093  
QY 1224 FVSTYDVFIDGNATIAKVTYDD-----TSKTSKYVYDVNVNDKTIETVTS-- 1269  
Db 1094 YDSEGRVLKQYSANSTEAPKVPETEOYSGHRLEKATNAKETYVYSYDADKKTLLMTQPN 1153  
QY 1270 -----DKLGVKTTTLTKTSANGNATKFSAADGALYKASDIATHLN 1311  
Db 1154 GRKVOGYNEAGNPQIVIDAEGLKITNTK-----YEGNVNVEDVD----- 1195  
QY 1312 TLAGDIQTAKGASQASSASYVDAGNKV--IYDSTDKKYQVNDKQGVDKNK-----E 1363  
Db 1196 --PNDVGTGK---ATESYOY-DKDGNTSVKDAYGTETYEYKNNNDVTMKDTEGNVTD 1248  
QY 1364 VAKDKLVAQAOTPDGTLAQMNVKSYNK-----EQVNDANKKQGINEDNAFTKLE---- 1414  
Db 1249 IAYDGLDAVSETDQS--GKSSAAVYDYKQNOIOSKSLASNTLTKDGSFEAKSGWNL 1306  
QY 1415 NAAKDTK-----TKNAAVTGDLNVAQAOTPLTEAGDTGTAKKLGELTITKG 1462  
Db 1307 TASKDRRKISVIADKSGVLGSKALEVLSQSISACTDHGYSATQTVBELENTTYTISGK 1366  
QY 1463 -QTDYTNK-----LTD-----NNTGVVAGTVDGFTVKLAKDLTINLSVNNAGTR 1503  
Db 1367 IKTDLAKSRAYFNIDRLDKQRIOWIHNEYSAIAGKNDWT---KROITFTTPANAKAV 1423  
QY 1504 I-----DEKGISEVD---ANGQAKANTPVLSANGLDLGKRSINIGAAVDD---- 1546  
Db 1424 VYMEYDHDKCKGKAWFEVQLEKGEVSSYNPVQSSFTSATENMNVSGASVDSBEGF 1483  
QY 1547 NDVAVNFKOFNEVAKTVNNLNQNSGASLPFW--TDANGKPIGTGKPKOKA--IKGAD 1602  
Db 1484 NDDVSLKAARTSASQAGSVTQOT-----VLQGSANDKPVYLTITGMSKASSVFTD 1535

QY 1603 GKYYHANANGVPVDKQKPIITDADKLANLAAGKPLDAGHOVVASLGGNSDAITLTNIKS 1662  
Db 1536 EKDYSLQAN-----VTYAD--GSTGIYNAKPPSGTQ-----EMRAAAVWIPKTK- 1577  
QY 1663 TLPOIDTPTNNGNAGQAQSLPSLSAAQSNAAV---KDVLENVGFNLQT-----NHQ 1713  
Db 1578 PINKVD-----ISILFKQSATGTWFDIDRLIEGSLLTAKSYDYSNGY 1620  
QY 1714 VDFVKAYDTVNFVNGTGADITSVRSAD-----GTMSNITVN---TALAAT 1755  
Db 1621 V--TKEDELGATSTYDYDETKKTSYSETDAKGEKTYTYDQADQLTNMTLSNGTSLHSY 1678  
QY 1756 DDDGNVL---IKAKDGKFKAD-DLM-----PNGSLKAGK-SASDAKTPGLSLVNP 1802  
Db 1679 DREGNEVSKTIRAGADQIVKFEYDVWGLVKTTPDLGNVLASEYDANSNLTKT---ISP 1734  
QY 1803 NAGKSGTG---DAV---ALNLSKAVFK-SKDGTMTTSSDGSIGQK---DNSS-ITL 1851  
Db 1735 NGNEVSLSYDGTDRVKSXSYNGTEKYIETDYKNGNETSVNKEQNTTKKRTFDKNRLTE 1794  
QY 1852 SKDGLNVGKV-----ISNVCKGTDKTDAAANVQOLNEVRNLLG----- 1889  
Db 1795 LTDR---GQSQTWTPSDSKLTKFSWIGHGQKGTNQFTYKNKLDQMIEMKDSYSSEFY 1851  
QY 1890 -----LGNAGNDNADGNQVNIADIKK--DPNSG-----SSNRTVTK--- 1924  
Db 1852 DENGNVQTFITGGGTSFSYDERNLVSSLHIGDKNGGDILTSEYVEDANGNRRTINSSA 1911  
QY 1925 AGTVL--CGKGNNDTEKATGGVQGVVDKGNANGDLSNVVWVKQKQSGKALLATYNAA 1982  
Db 1912 SKQVOYEYKGLNQLVKETHEDGTVIEYTDGFGN---RKTVTITKDGSSKTVNASENIM 1967  
QY 1983 GQTNLYLTNPAAEADIRINEQGI-----REFHVND-----GNOEPV- 2018  
Db 1968 NQ-----LTKVNDESISYDKNGNRFTSDGKFTYTDADENLNAVTKKEDKPFAT 2016  
QY 2019 ---QCRNGIDSSASGK-----HSVAIGFOAKADGAAVAITGROAGNQSIAIGDN 2066  
Db 2017 KYDEKGNRIQKTVNGKVTNYFYDGDLSNLVLETADNNV-----TKSYTYGDS 2065  
QY 2067 AQ-----ATGD-QSIAIGTGNVTVGKHSIGAIGDPSTVKA-----DNSYS 2104  
Db 2066 QQLLSYTYENGKKYFYHNAHGDIIAISDSCTKVAKYOYDAGWNPRTKTEASDEVKDNRYR 2125  
QY 2105 VG-----NNNQFI-----DATQTDVFGVGNNTV-----TE 2130  
Db 2126 YAGQYDEETGLYILMARYEPRNGVFLSLDPPDPSGDSLDQNGYAYGNNNPVNNVDPD 2185  
QY 2131 SNSVALGSNSAISACTHAGTQAKSDGTAGTTTATAGATVTKGFAGQATVAGVSVGASGA 2190  
Db 2186 GHWVWLNVNAGFAA--YDGYKAYKSG-----KGWKAAGAAAA----- 2220  
QY 2191 ERRIQNVAAGEVSATSDAVNGSOLYKATOGIANATNELDRIHON-----ENK 2239  
Db 2221 ----SNEGPKIEF-----CASPAYKFTKAVKITGHTRHGLNQSIGRNGGKGVNLR 2270  
QY 2240 ANAGISSAMAMASMPQA---YIPGRSMVTGGIATHNGQGAVALGSLKSLDNGQWFKING 2296  
Db 2271 LNA-VRSPKVKIKQPNGATKYVGKKATV---VLNKRKGVITAYGSSR-AGKSKHVFHTHG 2325  
QY 2297 SAD 2299  
Db 2326 KGN 2328

RESULT 13

PIP\_LACLC

ID PIP\_LACLC

AC PIP\_LACLC

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE PI-TYPE PROTEINASE PRECURSOR (EC 3.4.21.-) (WALL-ASSOCIATED SERINE

STANDARD;

PRT; 1902 AA.





QY 458 GTDGLKFTDUNSWTALEDTRITRKDKIGFSNKACTVDENKPYLDKOKLKVGNSTLNNGLT 517  
 Db 57 ATDTAATTQAIA-----TQLAAKGIDY-NLKNVQOQDIYVIVQMSAAPASENGTL- 110  
 QY 518 VNNTIGGSNKIQGVAGDGKIFADVNVSNAAKFGTTRITEEIGFADADGKV-----DKK 573  
 Db 111 --RTDYSSTAETQOEINKVIAQASV-----KAAVEQVTOQTAG--ESYGVVNGFSTK 160  
 QY 574 SPYLDKKQL-QVGGVKIT-----KDSGINA-----GDKISNVKDATD 610  
 Db 161 VRVVDIPKQLKQIAGVKTVTLAKVYVPTDAKANSMAVQWWSNYKYKGEGTVVVIDSGI 220  
 QY 611 DTDATVYKOLKQVQOQADGALOSFSIRDEKGOBEFTISNLSNCGTNTPTTITFACENGI 670  
 Db 221 DP---THKDMR-LUSDKDKVLKTSOV--BKFTDTAKHGRYFNSKVPYGFN-----YADNND 270  
 QY 671 SISNDIAKGVKVKYIDPINGLTTPKLTGSDKDKGTQOLVIEQVASGNDKNIIRGLSP-- 728  
 Db 271 TITDDT-----VDEQGHMVAGI-IGAN-----GTGDDPAKSVVGVAPEA 309  
 QY 729 ---TLPSINAGVRRTEOGNTITSDEDSKSAASIGDIINTGNFNKNSNSVGFVSTYNT 785  
 Db 310 QLLAMKVFTNSDTSATGTSATLVSATEDSAKIGA--DVLNMSLG-SDSGNQTLDEPELAA 366  
 QY 786 VDFIDGNATKATVYDETQTSKVTYDVNVDEKTIELTGDNKTKNIGYKVTTLTITTNAN 845  
 Db 367 VQANESGTAAVISAGNSGTSATGKNDYVGLQ---DNEMVGPFG--TSRGATTVAS 421  
 QY 846 GRATNESTTDNDALVNAKDIAENLNTLAKEIHTTKGTA-DTALQTEFKVKD---GATDDE 901  
 Db 422 -----AENTDVIITQAVITDGTGLQGPETIQLSSNDFGTSFQDK 461  
 QY 902 TITVKGDTQN---GKTVN-TLKLKGENGLT-----VATNKDG 935  
 Db 462 KFYVVKDASGLSKGVADYTDADAKGKIAIVKRGELTFADKOKYQAQAAGAGLIIVNNDG 521  
 QY 936 TV-----TFGINTQSLKAGDSTTLN-KDGLSIK----- 963  
 Db 522 TATPVTSMALTTFTTFFGLSSVTYQKLVDMVAHPDDSLGVLKIALTLVPNQKYTEDKMSD 581  
 QY 964 ---NPASNEQIQ--VGADGVKFAKVDKNSSTGIDGTSRITKDQIGFTGANGSL----- 1012  
 Db 582 FTSYGPVSNLSPKPDITAGGNINWSTONNGYTNMSTSWASP-----FIAGSOALLKQAL 637  
 QY 1013 -----DTPKPLHFKDLKVGVEITNTGINAGGKTIINIOGSDITQNSDVAIT- 1060  
 Db 638 NNKNPNFYAYYKOLKGTALTDFLKTVEMN-TAOPIN-----DI--NYYNVIVS 682  
 QY 1061 ---GGRVYDLKTELES-KINSAKTAQN-----SLHFSVADBO-----GNHFTVSNPY 1105  
 Db 683 PRQAGLVDMVKAADALEKNPSTVVAENGYPAVELKDFSTDKTKFLTFTNRTHELTY 742  
 QY 1106 SSYDTSKTSVITFAGE--NGI--TTKVNKVVRVVIDGT----- 1141  
 Db 743 QMDSNTDTNAVITYSATDPNSGVLYDKKIDGAAIKAGSNITVPAGKTAQIEFTLSLPKSED 802  
 QY 1142 -----KGLTPPKLTV-----GN-NNGKGVIVDSKDG--QNTITGLSNLTANVT 1181  
 Db 803 QOQFVEGFLNFKGSDGSRNLNLYPMGFQDWDNGK--IVDSLNGITYSPAGGNFGTVPLLT 860  
 QY 1182 NDGAGHALSQGLANDPDKTRAASIGDVLNAGFNLOQNGCAVDFVSYDIVDFIDGNATTA 1241  
 Db 861 NKWTGYGGMWYTDAGNQTV-----DDQAIAFSS-----DKNA--- 895  
 QY 1242 KVTYDDTSKTSKVYDVNVNDKTIETVSDKILGVKVTTLTKTSANGNATKFSAADGDAL- 1300  
 Db 896 --LYNDIS--MKYLLRLNSNVQVDIILDCQ--GNKVTTLIS-SSTNRKKYVYNAHSQOYIY 948  
 QY 1301 --VKASDIAHLNTLAGDIOTAKGASQASSASYVDADGNKVIYD---STDKKYIOV--- 1352  
 Db 949 YHAPAMD-GTYVDORDGNIKTADGGSYTYRISGVPEGGKQVDFVDFPKLDSKAPTVRHV 1007  
 QY 1353 -----NDKQGVKKNEKAKDKLVAQAQTPDGTTLAQMNVKSVINKEQVNDANKKOGINE 1405

RESULT 15  
 CNA STAAU  
 ID CNA STAAU  
 AC Q53654;  
 DT 15-DEC-1998 (Rel. 37, Created)

STANDARD;  
 PRT; 1183 AA.

Db 1008 ALSAKTENCKTQYLYLTAE-AKDDL-----SGLDATKSVKTAINEVTNLDATFTDAGTT 1059  
 QY 1406 DNAFIKGLENAADKTKNAAVTVGDLNVAQOTPLT-----FAGDTGTAKKLGCT---LT 1458  
 Db 1060 ADGYTK-IETPLSDEQ-----AQALGNGDNSAEIYLTDNASNATYDQDASVOKPOSTFDLI 1114  
 QY 1459 IKGQDGTNKLTDNNITGVVAGTDG-----FTVKLAKDLTNLSNVNAGGTRIDEKG--- 1508  
 Db 1115 VNGGGI-PDKISSTTTGYEANTOGGGTYTFESGYPAADV-----GTYTDAQGKKH 1163  
 QY 1509 ---ISFVDANGQAKANTPVLSANGLDLGCKRISNIGAAVDDNDNAVNFQKQNE-----VAKT 1561  
 Db 1164 DUMTYYDAATNSPTASMPVTNAD-----YAAQVLDLYADKAHTQLKLFHEDTTRVLTAPT 1216  
 QY 1562 VNNL--NNOSNCSASLPFVVT---DANGKPIGTGDKPKOKAIGKADGKYKHANANGVPV- 1615  
 Db 1217 FTDLKENNGSDQSEATIKVTGVSADTKTVNVGD-----TVAALDAQ--HFSVDVVPVN 1269  
 QY 1616 -----DKDG-----KPIT---DADKLANIAAHGKPLDAGHQVVASLGGSNDAI 1655  
 Db 1270 YGDNITIKVATDEDEGNTTTEQKTIITSSYDPMLKNSVT-----FDQG-----VTFGAN--- 1317  
 QY 1656 TLTNIKSTLPOIDTPNTGNANAGQAQSLPSLSAAQOSNAASVKDVLNVGNL----- 1707  
 Db 1318 ---EFNATSAKFYDPKGTGIATITGKVKHPTTTLQVDPKQIPIKDDLTFSFTLDGLTGQK 1374  
 QY 1708 -----QTNHINOVFKAYDTVNFVNGTGADITSV-RSADGTMSNITVTALAAATDDDG 1759  
 Db 1375 PFCVVVGDITQNTK---FQEAFTFILDVAVPTLSLSDSTDPAPYNNPNTQITGT--- 1426  
 QY 1760 NVLIKAKDKFYKADDLMPNGSLKAGKSASDAKPTTGLSLVNPNAKG---STGDAVAL- 1815  
 Db 1427 -----ATDNAQYLS--LSINGSSVASQYVD-----ININSGKPGHMAIDQPVKLL 1469  
 QY 1816 ---NNLSKAVFKSKDGTITTTVSSDGIISQCKDN-----SSITLSKQGLNV 1858  
 Db 1470 EGNVLTVAVTDESDNTTNNIT---VYIEPKKTLAAPTVPSTTEPAKTVTLTANSAAT 1526  
 QY 1859 GGRV--ISNVGKTKTDAAVQ--QLNEVRNLLGLGNAGNDNAGNOVNTADIKDOP--- 1912  
 Db 1527 GETVQVYSAOGKTYQDVPAAGVTVTANGTFKFKSTDLGYNESPAVDYV-VTNKADDDPAQ 1585  
 QY 1913 -NSGSSNRTVTKAGTVLGGKGNND-----TEKATG--GVQGVYDKDGNANGDL 1959  
 Db 1586 LQAAKQELNLIASAKTSLSAGKYDDATTTALAAATKQATALDQTNASVDSLGTGANRDL 1645  
 QY 1960 SNVVVTKQDKGSKALLATYNAAGQTNLYLTNNPAAEALDRINEQGIREFHVDNGQEPVQ 2019  
 Db 1646 QT-----AINQLAAKLPADKKTSL--NQLQSVKAALETDL-----GNQ----- 1682  
 QY 2020 GRNGIDSSAGSKHSAVIGFOAKADGEAAVAIGRQTOAGNOSIAIGDNAQATGDOQSTAGT 2079  
 Db 1683 ---TDSSTGK-----TFTAALDDLVA-----QAQACTQT---DQOLQATLAKVLDLAVL 1724  
 QY 2080 GNVVTKHSGAIGDPTSVKADNSYSVGNNNQFIDATOTDVFVGNNITVITSNVALGSN 2139  
 Db 1725 AKLAEG-----IKAAATPAEVGNK---DAATGKTWYADIADTLTSGQASADSD 1770  
 QY 2140 -----SAISACTHAGTQAKKSDGTAGTTTTAGATGT---VKGFAGQATYAGVAVS 2184  
 Db 1771 KLAHLQALOSLTKVAAAVEAAKTVGKGDTGTCTSDKGGQGTPTPTGPDIGDKGDEGS 1830  
 QY 2185 VGASGAERRIIONAAGEVSATST---DAVNGSOLYKATOGIANATFELDRH 2232  
 Db 1831 QPSSGG-----NIPTPNPTTTSTDDTTDRNCQLTSGKALPKTGETTER 1876

DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE COLLAGEN ADHESIN PRECURSOR.  
 GN CNA.  
 OS Staphylococcus aureus.  
 OC Bacillus; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Staphylococcus.  
 OX NCBI\_TaxID=1280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FDA 574;  
 RX MEDLINE=92165839; PubMed=1311320;  
 RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,  
 RA Lindberg M., Hoeck M.;  
 RA "Molecular characterization and expression of a gene encoding a  
 RT Staphylococcus aureus collagen adhesin.";  
 RL J. Biol. Chem. 267:4766-4772(1992).  
 RN [2]  
 RP ERRATUM.  
 RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,  
 RA Lindberg M., Hoeck M.;  
 RN J. Biol. Chem. 269:11672-11672(1994).  
 RN [3]  
 RP COLLAGEN-BINDING DOMAIN.  
 RC STRAIN=FDA 574;  
 RX MEDLINE=94032261; PubMed=8218209;  
 RA Patti J.M., Boles J.O., Hoeck M.;  
 RA "Identification and biochemical characterization of the ligand  
 RT binding domain of the collagen adhesin from Staphylococcus aureus.";  
 RL Biochemistry 32:11428-11435(1993).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.  
 RX MEDLINE=97475225; PubMed=9334749;  
 RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,  
 RA Moore D., Jin L., Schneider A., Delucas L.J., Hoeck M.,  
 RA Narayana S.V.L.;  
 RT "Structure of the collagen-binding domain from a Staphylococcus  
 RT aureus adhesin.";  
 RL Nat. Struct. Biol. 4:833-838(1997).  
 CC -1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO  
 CC COLLAGEN-CONTAINING SUBSTRATA.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. CELL WALL.  
 CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS  
 CC IN THE REGION OF THE MEMBRANE ANCHOR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions. There are no restrictions on its  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M81736; AAA20874.1; .  
 DR PDB; 1AMX; 24-JUN-98.  
 DR InterPro; IPR001899; .  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; FALSE\_NEG.  
 KW Signal, Repeat; Transmembrane; Cell wall; 3D-structure.  
 FT SIGNAL 1 29  
 FT CHAIN 1 29  
 FT DOMAIN 30 1183  
 FT TRANSMEM 30 1157  
 FT DOMAIN 1158 1177  
 FT DOMAIN 1178 1183  
 FT DOMAIN 151 318  
 FT DOMAIN 533 1093  
 FT DOMAIN 1093 1157  
 FT DOMAIN 1151 1156  
 FT REPEAT 533 719  
 FT REPEAT 720 906  
 FT REPEAT 907 1093  
 FT REPEAT 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;  
 SQ SEQUENCE

Query Match 3.2%; Score 379; DB 1; Length 1183;  
 Best Local Similarity 21.1%; Pred. No. 7.9e-07;  
 Matches 309; Conservative 182; Mismatches 543; Indels 430; Gaps 71;  
 QY 289 NNNKAYVEG-NGSNIKSSKATGNGLFSGSSTIKRKINVGAGYEDTDVAVNAOLKA-- 345  
 DB 23 NKNEAFAARDISSTNVTDLTVSPSKIEDGGTKTVKMTFDDKNGKIQNGDMIKVAMPTSGT 82  
 QY 346 --VENLAKR-QITFKGDONGTGKKKLGTLTIKGETQADKLTDDNNIGVVTDNNTGLK 402  
 DB 83 VKIEGYSKTVPLTVKGEQGVQAVITPDGATITF---NDKVKLSVSGF-----AEFE 132  
 QY 403 VKLANLSLETVSTKNLTASEKV--TVSGNNTAELOSGGLTFTPTTNASTDKTVYGTG 461  
 DB 133 VQ-GRNL-----TQTNTSDDKVATITSGNK-----STNVTVHKSAGTSS 171  
 QY 462 LKFTDNSNTALEDTRITKDKIGESNAGTVDENKPYLDKDKLVKGNSTLNGGLTVNNT 521  
 DB 172 VFYYKTGDMLPEDTHVR----WFLN-----INNEKSYVSKD-----ITIKD- 209  
 QY 522 IGGSNKQIOVGADGKIFADVNVNYSNAK---FGTTRITEEIGFADADGKVDKSPYLD 578  
 DB 210 -----QIQ-GGQQLDLSTLNINVTGTHSNYSGSALTDFEKAAPGSKTVTDNKTATID 262  
 QY 579 KKQLQGVGVKITKDSGINAGDOKISNVKDATDDTDVATYKQLKOVQDADGALQSFSIRD 638  
 DB 263 -----VTIPOGCGYSNFSINFK-----TKITNEQKQEFVNNSQAWYQEHGKKE 306  
 QY 639 EKGQEF--TISNLSNGNTPTNPTFETITFAGENGISISNDIAKGVKVGIDPINGLEITPKL 696  
 DB 307 VNGKSFNHTVHINANAGIEGT-----VKELKV----- 335  
 QY 697 TVGSDKDGKQLVIEQVAVSGNDKNTIIRGLSPITLPSITNAGV-----RTTE 743  
 DB 336 -LQDKDKOTKAPIANVKFKLSKDGSVVKNQKQKEIITDANGIANIKALPSGDYILKEIE 394  
 QY 744 QGNTITSDKSKAASIGDILNTGFNLKNNNSVGFVSTVNTVDFIDGNATKATYDET 803  
 DB 395 APREYTFDKDKYFPFTMKDNDQY-----FTTIE-----NAKATEKTDYS 436  
 QY 804 NQ-----TSKV-----TYDVNVDEKTIELTDGNGKTKIGVKTTTLTTN-----ANG 846  
 DB 437 AQKWEQTKVPTIYFKLYKODDQNTTPV--DKAEIKKLEDDGTTKVTWSNLPEKNDNG 494  
 QY 847 KATNFSTDDNALVNAKDAENLTLAKEIHTTKGADTALQTFKVKDKCATDDEITVVG 906  
 DB 495 KAIKY-----LVKEVNAQ----- 508  
 QY 907 KGTQNGKVTNLLKKGNGLTVAENKDGTVTFGINTQSLKAGDSTTLNKDGLSIKNPA 966  
 DB 509 EDTTPEGYT-----KKNGLVVTNTEKPIETTSI---SGEKVWDDKO-NQDG---KRP- 554  
 QY 967 SNEQIOVG--ADGVKFAKVDKNGSSTGIDGTSRIK----DOIGFTGANGSLDTPKPHLT 1020  
 DB 555 --EKVSYNLLANGKVKTLDTVSETNWKYEFKDLPKYDEGKIEYT-----VT 600  
 QY 1021 KKKLVGEVEITNTGINAGKKKTIQSGDITQN--SNDAVTGGRVYDLKTELESKINS 1078  
 DB 601 EDHVKDYTTDINGTTIT--NKYTPGTSATVTKNWDNNODGKRPTEIKVELVDGKAT 658  
 QY 1079 AKTA-----QNSLHEFSVADEQGNHFTVSNPYSYDTSKTSVITFAGENGITTKVNGKV 1133  
 DB 659 GKTAILNESNNHTWTGLDEKAKGQQVK--YTVEELTKV-----KGYTHVDNN- 706  
 QY 1134 VRVGDQTKGLTTPKLTGVNNNGK-IVIDSQDQNTITGLSNTLANVYNDG-AHALSQ 1191  
 DB 707 -----DMGNLIVTNKYTPETTSISGEKVVDDKDNQ-----DGKPEKVSV 746  
 QY 1192 GLANDTDKTRAASIGDVLNAGFNLQ-----NGBEAVDFVSTYDTV-DF---IDGNATTAK 1242  
 DB 747 NLLADGKVKTLDTVSETNWKYEFKDLPKYDEGKIEYTVTDEHVKDYTTDINGTTITNK 806

```

QY 1243 VTYYDSKTSKVYD--VVDNK-----TIEVTSDKKLGKVTTLTKTS----- 1284
Db 807 YTPGETSATVTKNWDNNQDGRPTFKVELYODGKATGKTKAILNESNNWHTWTGLDE 866
QY 1285 -ANGNATKESAADGDAIVRASDIATHL--NTLAGDIOTAKGASQASSASYVDADGNKVI 1341
Db 867 KAGQOVKTV---EELTKVGYTHVDNNDGNLIVTNKYTPETTS-----ISGEKV- 916
QY 1342 YDSTDKKYOVNDKGOVDNKNKEVAKDLVA---OQTPDGTLAOMNVKSVINKEQVNDAN 1398
Db 917 WDDKD-----NQDGRPEKVSVNLLANGKVKTLDT-SETNWKYEFKDLPKYDEG 966
QY 1399 KK--QGINEDNAFIKLENAKDTKTNAAVTVGDLNVAQAOTPLTFAGDTGTTAKKLGET 1456
Db 967 KLEYTVTDEH-----VKDYTT-----DINGTTITNKYTPGETSATVTKNWD 1009
QY 1457 LTIKGGQTDITNKLTDNNIGVAGTDFGVKLAKDITNLNSVNAAGTRIDEKGISFYDANG 1516
Db 1010 NNDGKRPT-----EIKVELYODGKATGKTKAILNESNNWHTWTGLDEK-----AKG 1057
QY 1517 -QAKANTPVLANSGLDGGKRISNIGAAVDDNDVAFKQFNEVAKTVNNLNQSNQSGASL 1575
Db 1058 QQVKYTVDEL-----KNGYTHVDNNDMGN----- 1084
QY 1576 PFVVTDANGKPIGTGKPKOKAI-----KGADGKYIHAN-ANGVVPVDKDKPITDADK 1627
Db 1085 -LIVTN-----KYTPKPKPKPIYEPKPKDKPTTPKPHSNKVKPTPPDKPSKVKRDKDQ 1137
QY 1628 LANLAHAGKPL-----DAGHQVVAS 1647
Db 1138 KDKTKPENPLKELPKTGKIIITS 1161

```

Search completed: September 13, 2001, 12:54:29  
Job time: 899 sec

**THIS PAGE BLANK (USPTO)**